DOCUMENT RESUME

ED 225 827 SE 040 186

AUTHOR	Herbert, Martin
TITLE	Fourth Grade MANS Test Data. Evaluation Report 5-B-1. Extended Pilot Trial of the Comprehensive School Mathematics Program.
INSTITUTION	CEMREL, Inc., St. Louis, Mo.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Nov 78
NOTE	144p.; For related documents, see SE 040 181-196 and SE 040 340-348.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Educational Research; *Elementary School Mathematics; Evaluation Methods; *Grade 4; Intermediate Grades; Mathematics Achievement; *Mathematics Curriculum; Mathematics Instruction; Program Effectiveness; *Quantitative Tests
IDENTIFIERS	*Comprehensive School Mathematics Program; Mathematics Applied to Novel Situations Test; *Mathematics Education Research

ABSTRACT

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and was funded by the National Institute of Education (NIE). Its major purpose is the development of curriculum materials for the elementary grades. Beginning in 1973, CSMP ran an extended pilot trial of its elementary program. This report describes the results of administering a series of tests to nine fourth-grade CSMP classes in the St. Louis area and to eight comparison classes. The tests are collectively called the Mathematics Applied to Novel Situations (MANS) Test. The study was conducted in May 1978, at the end of a school year that constituted the first of 2 years of the Extended Pilot Trial (EPT) of CSMP fourth-grade materials. Results indicated that CSMP pupils did seem to be consistently better at certain kinds of tasks than non-CSMP students. (MP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED225827

EXTENDED PILOT TRIALS OF THE COMPREHENSIVE SCHOOL MATHEMATICS PROGRAM:

EVALUATION REPORT SERIES

Evaluation Report 5-B-1
Fourth Grade MANS Test Data

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

CEMREL INC.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SE040186

Extended Pilot Trial of the Comprehensive School Mathematics Program

Evaluation Report 5-B-1
Fourth Grade MANS Test Data

Martin Herbert

November, 1978

Developed by CEMREL, Inc., a private nonprofit corporation supported in part as an educational laboratory by funds from the National Institute of Education, Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement should be inferred.

Copyright on these materials is claimed only during the period of development, test, and evaluation, unless additional authorization is granted by the National Institute of Education, to claim copyright on the final materials. For information on the status of the copyright claim, contact either the copyright proprietor or the National Institute of Education.

Description of Evaluation Report Series

The Comprehensive School Mathematics Program (CSMP) is a program of CEMREL, Inc., one of the national educational laboratories, and is funded by the National Institute of Education. Its major purpose is the development of curriculum materials for grades K-6.

Beginning in September, 1973, CSMP began an extended pilot trial of its Elementary Program. The pilot trial is longitudinal in nature; students who began using CSMP materials in kindergarten or first grade in 1973-74, were able to use them in first and second grades respectively in 1974-75, and so on in subsequent years. Hence the adjective "extended".

The evaluation of the program in this extended pilot trial is intended to be reasonably comprehensive and to supply information desired by a wide variety of audiences. For that reason the reports in this series are reasonably non-technical and do not attempt to widely explore some of the related research issues. The list of reports from the first two years of the extended pilot trial is given on the next page. The most comprehensive of these are the following:

1-A-1: Overview, Design and Instrumentation
1-A-3: Final Summary Report, Year 1
2-A-1: Final Summary Report, Year 2
3-B-1: Summary of Second and Third Grade Test Data Year 3
4-A-1: Final Summary Report, Year 4
and 5-A-1: Summary Data Through Year 5

Longitudinal Pilot Study of the
Comprehensive School Mathematics Program

Evaluation Report Series

Evaluation Report 1-A-1	Overview, Design and Instrumentation
Evaluation Report 1-A-2	External Review of CSMP Materials
Evaluation Report 1-A-3	Final Summary Report Year 1
Evaluation Report 1-B-1	Mid-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-2	End-of-Year Test Data: CSMP First Grade Content
Evaluation Report 1-B-3	End-of-Year Test Data: Standard First Grade Content
Evaluation Report 1-B-4	End-of-Year Test Data: CSMP Kindergarten Content
Evaluation Report 1-B-5	Test Data on Some General Cognitive Skills Related to CSMP Content
Evaluation Report 1-B-6	Summary Test Data: Detroit Schools
Evaluation Report 1-C-1	Teacher Training Report
Evaluation Report 1-C-2	Observations of CSMP First Grade Classes
Evaluation Report 1-C-3	Mid-Year Data from Teacher Questionnaires
Evaluation Report 1-C-4	End-of-Year Data from Teacher Questionnaires
Evaluation Report 1-C-5	Interviews with CSMP Kindergarten Teachers
Evaluation Report 1-C-6	Analysis of Teacher Logs
Evaluation Report 2-A-1	Final Summary Report Year 2
Evaluation Report 2-B-1	Second Grade Test Data
Evaluation Report 2-B-2	Readministration of First Grade Test Items
Evaluation Report 2-B-3	Student Interviews
Evaluation Report 2-C-1	Teacher Questionnaire Data
Evaluation Report 2-C-2	Teacher Interviews, Second Grade
Evaluation Report 2-C-3	Teacher Interviews, First Grade
Evaluation Report 3-B-1	Second and Third Grade Test Data Year 3
Evaluation Report 3-C-1	Teacher Questionnaire Data Year 3
Evaluation Report 4-A-1	Final Summary Report Year 4
Evaluation Report 4-B-1	Standardized Test Data, Third Grade
Evaluation Report 4-B-2	Mathematics Applied to Novel Situations (MANS) Test Data
Evaluation Report 4-B-3	Individually Administered Problems, Third Grade
Evaluation Report 4-C-1	Teacher Questionnaire Data, Third Grade
Evaluation Report 5-A-1	Summary Data Through Year 5
Evaluation Report 5-B-1	Fourth Grade MANS Test Data
Evaluation Report 5-B-2	Individually Administered Problems, Fourth Grade
Evaluation Report 5-C-1	Teacher Questionnaire and Interview Data, Fourth Grade

Key to Indexing

Evaluation Reports are labelled m-X-n,

where m is the year of the pilot study, with 1973-74 as Year 1.
X is the type of data being reported where A is for overviews and summaries, B is for student outcomes and C is for other data.
n is the number within a given year and type of data.

Table of Contents

Introduction 1

Setting 3

Comparison of Class Mean Scores 8

Other Findings 18

Summary 31

Appendix A: Description of MANS Scales A1

Appendix B: Scale-by-Scale Statistical Analyses B1

Introduction

In the fall of 1973, the Comprehensive School Mathematics Program (CSMP) began a longitudinal pilot study of its Elementary School Program. Over 100 teachers began using the program, either in first grade or kindergarten. In subsequent years some of these classes continued with CSMP materials in the next grade levels, and many new classes began using CSMP materials. The 1977-78 school year constituted the first of two years of the Extended Pilot Trial (EPT) of CSMP fourth grade materials.

This report describes the results of administering a series of tests to the 9 fourth grade CSMP classes in the St. Louis area and to 8 comparison classes. The tests are collectively called the MANS Test (Mathematics Applied to Novel Situations, although some of the test scales consist of ordinary computation problems), and are intended to compare CSMP and Non-CSMP students on a variety of mostly novel problem situations. It is hoped that many of the thinking skills underlying the scales reflect those embodied in many of the CSMP lessons and hence one hopes these scales can be viewed as transfer tasks.

Previous Evaluation Reports describe the administration of MANS Tests at lower grade levels; third grade (Evaluation Report 4-B-2) and second grade (Evaluation Report 3-B-1). Another report, 5-B-2, describes the results of a series of individually administered tasks, similar in nature to the MANS, which were given to samples of students from these classes.

Setting

Selection of Classes

The 1977-78 school year was the first year of the Extended Pilot Test of fourth grade CSMP materials (Mathematics for the Intermediate Grades, Parts I and II). Altogether there were 36 such classes. (This figure excludes 13 fourth grade classes which were new to CSMP and were studying a special "entry" program; performance of entry classes will be examined next year in the second year of the fourth grade EPT.)

These 36 classes were located in 16 sites. The St. Louis area was the only site in which there were sufficient CSMP and comparison classes to make a study of this kind worthwhile in 1977-78. In the St. Louis area there were 10 veteran (i.e., non-"entry") fourth grade classes, located in the Normandy, Ferguson-Florissant and Ladue school districts. One of these classes stopped using CSMP midway through the year. The remaining 9 were compared with 8 Non-CSMP classes, selected from adjacent schools where possible, but in any case with the intent that this group of students, teachers and schools be as similar to the CSMP group as possible.

Early in the year the Kuhlmann-Anderson test, a test of general mental ability, was administered to students in these 17 classes. Thus it was possible to compute the mean score for each class, and to use this mean as a covariate in the subsequent analysis of class means. It was also then possible to compare CSMP and Non-CSMP students at different levels of ability, as measured by the Kuhlmann-Anderson.[1]

[1]The score used in all cases was the total number correct on the Kuhlmann-Anderson excluding the 2 math sections. The math sections were excluded because it was desired to investigate the cumulative effects of CSMP through fourth grade, rather than merely the effect of the fourth grade curriculum by itself.

Only those students who were present when the Kuhlmann-Anderson Test was administered were included in the analyses of the various test results. This excluded those students who happened to be absent that day or who joined the class after that date (altogether an average of about 2½ students per class). The average number of students per class was about 22 for both CSMP and Non-CSMP classes.

The mean KA score across CSMP and Non-CSMP students turned out to be fairly close: 79.9 for CSMP classes versus 77.9 for Non-CSMP, which is a difference of only about an eighth of a standard deviation. The distribution of class means was somewhat different between CSMP and Non-CSMP classes. This is perhaps best illustrated by Table 1 below, which lists the approximate percentile ranks of each class.

Table 1
Percentile Ranks, Kuhlmann-Anderson Test

CSMP Classes	Non-CSMP Classes
81	78
78	73
75	71
73	71
71	69
66	63
52	60
48	34
48	

Table 1 shows that the corresponding percentile ranks are fairly close, except for the lower scoring classes. There are no CSMP classes that match the lowest scoring Non-CSMP class. Table 1 also shows something about the overall ability level of the whole set of 17 classes. The median percentile rank was about 70 for both CSMP and Non-CSMP classes. This also turned out to be the case at the individual level; about half the students in this study had percentile ranks above 70. Thus this group of classes was well above average in ability.

The MANS Test

The MANS tests are specially developed test scales designed for the purpose of comparing CSMP and Non-CSMP students' ability to deal with certain kinds of novel mathematical situations. The scales are intended to be related to some of the thinking processes which seem to occur in the CSMP curriculum, but without containing the language or content of the curriculum.

Previous MANS Tests were developed for use at second and third grade levels, and the present ones build on those previous efforts. The development of the various scales is a fairly lengthy process and includes extensive reviews (one group of reviewers being the external CSMP Evaluation Panel) and pilot tests. Appendix A of Evaluation Report 4-B-3 describes in detail how this process worked with the third grade MANS tests; the process was similar for the present scales.

Altogether, 25 scales were finally included. About half of these were at least based on, if not very similar to, previously used scales. The typical scale consisted of 5-9 items, required about 5 or 6 minutes (though there was usually no fixed time limit and was preceded by directions which varied from a few seconds to over three minutes (the latter being "mini-lessons" for those tasks which were particularly novel or unusual).

This year a number of these scales were not of the usual type; they were not novel in either format or substance. Because of a desire to keep the total testing time to a minimum and to use school-administered test data whenever possible, it was decided to use the standardized mathematics tests administered by the respective school districts as the basis for making CSMP/Non-CSMP comparisons on the achievement of what might be called the standard skills and concepts. But since each district used a

different test for this purpose it would not be possible to compare the whole set of classes on common items and scales. Thus some scales were created to cover the most basic and widely taught content from fourth grade. An analysis was made of 6 of the leading standardized tests and of some of the most common mathematics textbooks. In order to reflect what seemed to be the most heavily emphasized topics, scales were created in computation (Scales #1, 2, 3, 4), word problems (#5, 6, 8), fractions (#11), measuring (#12), and place value (#21).

Two test booklets, X and Y, were created. Each required a total of 50-60 minutes to be administered by a specially trained tester (one person did all the testing). Within each booklet the format was the same. For each scale the tester explained what the required task was and did one or two samples with the whole class, all according to a carefully designed script. Then the students went ahead and did the test items on that scale, which involved one or more pages. They did not turn the page to begin a new scale until they were told to do so by the tester - then everyone turned at once - and the same procedure was followed in the next scale.

Thus for any given scale, each student in the class had the same amount of time. For most scales this time limit was fairly flexible, with a suggested time limit for the tester, which could be extended if students were still working . For some scales which involved estimation, and in which laborious calculations were discouraged, strict time limits were adhered to.

Each of Test Booklets X and Y had two forms which differed only in 2 or 3 pages. For example, the fourth page of one form of Test X was the Computation-Multiplication Scale; the fourth page on the other form of Test X was the Computation-Division Scale. A random half of each class got each form. This was possible because the directions were the same for each (essentially "Do the problems on this page") and about the same

amounts of time were required for each. Altogether there were 5 such pairs of "matched" scales; for these 10 scales, all statistics are therefore based on half the number of students within each class. The first page of Appendix A lists all the test scales, showing the number of items, the approximate time required, which pairs of tests were alternated, and where results for that scale can be found. The remainder of Appendix A shows, for each scale, reduced copies of the relevant page(s) from the test booklet and a summary of the directions.

Comparison of Class Mean Scores

In order to simplify the reporting somewhat, similar test scales will be grouped together, described briefly and the results given. This grouping is somewhat arbitrary, but should make it easier for the reader to keep track of the various scales.

Appendix B gives data from the complete set of scales, one after the other, with two pages devoted to each scale. The test items are listed, together with a great deal of statistical data including item analyses, graphs of class means and Analysis of Covariance data. In fact Appendix B by itself could constitute a rather full report of the MANS testing.

Computation Scales

There were four computation scales, one in each of the basic operations, containing from 5 to 7 test items each. The format was mixed between horizontal and vertical and the difficulty ranged from basic facts to algorithms (simple two digit multiplication and one-digit division).

Table 2

Summary Data, "Computation" Scales

Scale	Number Of Items	Adjusted Means [1]		p-value from ANOCOVA F-Test p less than: [2]
		CSMP	Non-CSMP	
#1 Computation- Addition	5	4.4	4.0	.11
#2 " Subtraction	5	3.6	3.5	.64
#3 - Multiplication	7	5.7	5.3	.09
#4 " - Division	6	3.3	3.5	.54
Total	23	17.0	16.3	

[1] These are class means adjusted for differences in ability as measured by the Kuhlmann-Anderson score; CSMP classes had slightly higher KA scores so their MANS scores would be adjusted downward while the Non-CSMP scores would be adjusted upward. However, for the entire 25 MANS scales, these adjustments were almost always less than .1 and never as much as .2

[2] This is the p-value of the resulting F test from the Analysis of Covariance on the class means with 1 and 14 degrees of freedom. Low p-values are sometimes said to be "significant" if they are less than say .05; i.e., there is a 5% possibility that the observed differences between CSMP and Non-CSMP classes could have happened merely by chance had there been no differential curriculum effects.

It can be seen that there were no significant differences between CSMP and Non-CSMP classes on any of these scales. CSMP classes did slightly better in addition and multiplication; Non-CSMP classes did slightly better in division. This reflects exactly the results of a standardized test administered last year in third grade (see Evaluation Report 4-A-1), except that then more items were used and the differences on these 3 types of items were significant.

Scales 1-3, particularly Addition, were easy for most students. This possible ceiling effect may help explain why the Computation scales were, of all the MANS scales, the least predictable from Kuhlmann-Anderson mean scores. Although moderate correlations were still attained, class mean scores on these scales depended less on ability level of the class than did all the other scales. (Compare the graphs of class means in Appendix B for Addition (page B5) and Division (page B11) with those of the other MANS scales.)

Word Problems

There were four scales dealing with Word Problems. The results are shown in Table 3, below.

Table 3
Summary Data
"Word Problem" Scales

Scale	Number Of Items	Adjusted Means		p-value from ANOCOVA F-Test p less than:
		CSMP	Non-CSMP	
#5 Word Problems - One Step	8	6.2	5.9	.43
#6 " " - Two Step	6	4.1	3.1	.01
#7 " " - Hits & Misses	5	1.8	1.1	.06
#8 " " - Miscellaneous	6	2.7	2.2	.23
Total	25	14.8	12.3	

CSMP classes did significantly better on Scale #6, Two-Step Word Problems. A fairly typical item of this scale was the following:

> There are 40 apples in my barrel now.
> I will eat 2 apples every day.
> How many apples will be left in my barrel after 5 days.

This particular item required either repeated subtraction, or multiplication and then subtraction. Other items required different combinations of operations.

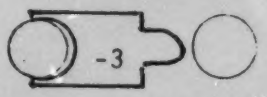
The difference in favor of CSMP classes on Scale #7, Hits and Misses, almost reached the .05 level. This scale concerned the calculation of game scores in which "hits" were worth 5 points while "misses" lost 3 points. It is worth noting, however, that this was a very difficult scale; almost half the classes had a mean score of less than 1 out of the 5 items correct and about half the students didn't get any items correct. The graph of class means (see page B17) also reflects a rather irregular pattern of scores. It may be that many students were unable to figure out how the game was played.

The non-significant difference in favor of CSMP classes on Scale #8, Miscellaneous, was due almost entirely to their superior performance on the two items which have to do with decimals (dollars and cents, and an automobile odometer).

Number Relationships

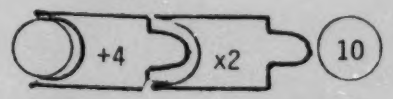
The 5 scales in this group were Scale 18, Solving Number Machines; Scale 19, Using Number Machines; and Scales 20a), 20b) and 20c) called collectively "Boxes". The description of these scales will be necessarily brief here, but each one was accompanied by considerable explanation by the tester and practice examples for the students (see Appendix A).

Scale 18, Solving Number Machines, had to do with machines like the following:



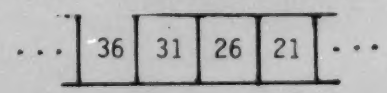
The left hand circle is always for the number we started with and the right hand circle is for the number we ended up with after the machine had done its job; in this case subtracting 3. For each problem of Scale 18, an unlabelled machine was shown in operation with various pairs of inputs and outputs, with one of the inputs or outputs missing. The task was to figure out from the given information what the machine was doing each time, and then to figure out the missing input or output.

Scale 19, Using Number Machines, also used machines, but they were always labelled and usually hooked up with other machines. For example:

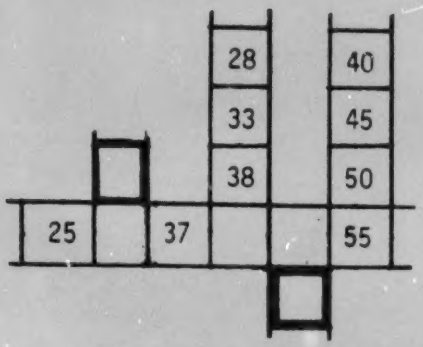


The task was to figure out the missing input (in this case) or output.

The three scales 20a, 20b, 20c, all dealt with "boxes", as in the "squares" of a crossword puzzle. For any given problem, the numbers are always increasing or decreasing by the same fixed amount when one goes horizontally and by some different amount when one goes vertically. A typical item from Scale 20a) asked whether 86 would ever be in one of the boxes (which extend both ways) of:



A typical item from Scale 20b) asked the students to fill in the heavy boxes in the following diagram, having been told that horizontal counting is by 6 and vertical counting is by 5.



Scale 20c) also required completion of boxes except that instead of "counting by" there was "multiplying by".

Table 4
"Number Relationships" Scales

Scale	Number Of Items	Adjusted Means		p-value from ANOCOVA F-Test p less than:
		CSMP	Non-CSMP	
#18 Solving Number Machines	8	6.1	5.2	.02
#19 Using " "	9	6.5	4.9	.01
#20a) Boxes- 1	9	6.9	6.0	.01
#20b) Boxes- 2	9	5.9	5.3	.03
#20c) Boxes- 3	7	3.6	1.6	.01
Total	42	29.0	23.0	

The differences in favor of CSMP classes on these scales are large and consistent and provide very strong evidence that CSMP students do learn to "think relationally" at least insofar as that ability is reflected in these 5 scales. Scales 18 and 19 are revised versions of scales used last year in third grade, when similar significant results were obtained.

Estimation

All 5 scales in this group are based on, and in some cases identical to, scales used last year. All involved calculations which, in one way or another, required only approximate solutions. In fact, exact calculations were not possible in the 1½-2½ minutes allowed for these various scales (10-15 seconds per item). These tasks required thoughtful but quick decisions.

In each of Scales 14a), 14b) and 14c) a series of computation problems was given and the task was to determine which of several given intervals contained the answer to the problem. For example, which interval contains the answer to 2 x 49?

(0,10) (10,50) (50,100) (100,500) (500,1000)

One doesn't need to know that 2 x 49 = 98, merely that it's less than 2 x 50 = 100

Scale #16, Multiplying by 2, 5 or 10, posed several questions like the following:

"100 is about ____ times as large as 19"

where the blank was to be filled in by whichever of 2, 5 or 10 was best.

Scale #17, Which is Larger, gave a series of pairs of computation problems and in each pair the student had to select the problem which had the largest answer, for example, 38 x 38 versus 39 x 31. Altogether there were 12 test items, 4 each dealing with multiplication, division and fractions.

Table 5
Summary Data
"Estimation" Scales

Scale	Number Of Items	Adjusted Means		p-value from ANOCOVA F-Test p less than:
		CSMP	Non-CSMP	
#14a) Estimation-Addition	8	6.3	6.0	.25
#14b) " Multiplication	7	5.9	5.3	.01
#14c) " Division	7	3.7	3.3	.03
#16 Multiplying by 2, 5 or 10	13	8.8	8.6	.55
#17 Which is Larger	12	7.5	6.5	.04
Total	47	32.2	29.7	

The results from Scales 14a) and b) reflected exactly the results from last year when these same scales were administered to third graders, namely, no significant difference on the addition items and a significant difference in favor of CSMP on the multiplication items. (A subtraction scale was also administered to fourth graders in an individualized setting and on this scale also, the results agreed with third grade data, namely no significant difference; see Evaluation Report 5-B-2.) CSMP students did signifcantly better on the division items of Scale 14c), a scale which was not administered last year.

In summarizing the results of the estimation tasks of Scale 14 then, one can say that CSMP students do better when the operation is multiplication or division and there is no difference when the operation is addition or subtraction.

There is a difference however for Scale #16 from last year to this year. The first 10 items were also administered last year and produced mean scores of 6.3 and 5.6 last year (a significant difference) versus 7.4 and 7.1 this year (a non-significant difference). Thus the CSMP advantage from last year appears to be disappearing.

On Scale #17, the same format was used last year, but most of the items were new. The CSMP advantage on this scale turned out to be mostly because of better performance on items involving division with the "÷" sign.

Large Number Computation

These two scales were composed of problems given in an open sentence format illustrated by two examples below.

$\square \times 4 = 400$

$800 \div \square = 200$

The problems in Scale 9 contained either addition or division signs and the problems in Scale 10 contained either subtraction or multiplication signs. Of course the process used by the student may or may not correspond to the written sign, as in the first example above where the student might ask either "4 times what equals 400?" or "What is 400 ÷ 4?"

The numbers were easy to work with, usually multiples of one hundred or fifty.

Table 6

Summary Data

Large Number Computation

Scale	Number Of Items	Adjusted Means		p-value from ANOCOVA F-Test p less than:
		CSMP	Non-CSMP	
#9 Large Number Computation (+, ÷)	9	4.9	3.5	.01
#10 Large Number Computation (-, x)	8	5.2	4.2	.01
Total	17	10.1	7.7	

The significant differences in favor of CSMP classes is consistent with previous data on these kinds of tasks.

Other Scales

The remaining 5 scales did not fit into one of the previously described groups of scales.

#11 Fractions: There were two kinds of items. Four items required taking $\frac{1}{2}$ or $\frac{1}{3}$ of a number; the other four items required addition or subtraction of fractions with like denominators.

#12 Measuring: In each item of this scale, a rod was shown aligned in some way with a ruler (not always starting at the zero mark). The task was to determine how long the rod was, various rulers being marked in halves, quarters and tenths of an inch. For example:

#13 Using Hints: Three very large addition and two multiplication problems were given, complete with answer. In each problem, one digit of one of the numbers in the problem was crossed out and a new one written above. The task was to figure out what the answer to the problem would now be, given this change. This was a matter of making the appropriate adjustment in the original answer. Not enough time was given to get the new answer from scratch.

#15 Labelling Number Lines: The items were of the following type:

#21 Place Value: In the items of this scale, a number line marked in either 10's, 100's, or 1000's was given and the task was to either indicate with an x on the number line approximately where a given number was located or name the approximate number located by a given x. For example:

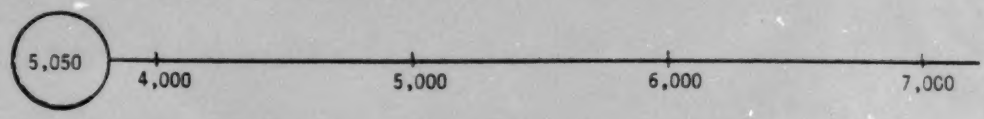


Table 7
Summary Data
Other Scales

Scale	Number Of Items	Adjusted Means		p-value from ANOCOVA F-Test p less than:
		CSMP	Non-CSMP	
#11 Fractions	8	4.3	3.5	.05
#12 Measuring	6	3.2	2.7	.05
#13 Using Hints	5	1.9	1.4	.10
#15 Number Line Labelling	6	3.8	2.8	.01
#21 Place Value	8	5.0	4.9	.65

In each of the three scales on which there was a significant difference in favor of CSMP, it is possible to partition the items into 2 distinct types. On one type CSMP students did better and these items accounted for the differences each time; on the other type of items there were virtually no differences. The items which made the difference were:

Scale #11 - taking 1/nth of a whole number (as opposed to adding or subtraction fractions of like denominators).

Scale #12 - measuring involving tenths of an inch (as opposed to halves or quarters).

Scale #15 - marking points outside an interval defined by two marked points as in the example on the previous page - i.e., extrapolation (as opposed to within the interval - i.e., interpolation).

Other Findings

Scores by Ability Level

Based on Kuhlmann Anderson scores, all students were assigned to one of four ability groups in such a way that there were approximately 35-40 CSMP and Non-CSMP students in each group. Because of the overall above average-ability level of the group of 17 classes, the division between the lowest quarter (Q_1) and the next lowest quarter (Q_2) corresponded not to the 25th percentile on the test, but to about the 50th percentile. Similarly the other divisions occurred at about 70th and 80th (versus 50th and 75th if these students had been perfectly representative of all fourth graders).

Within each ability group, mean scores were derived on each MANS test and on the Kuhlmann Anderson Test. Figure 1, below, shows a graph of the total of these means across all the MANS scores for each ability group plotted against the Kuhlmann Anderson, this being done for both CSMP and Non-CSMP students.

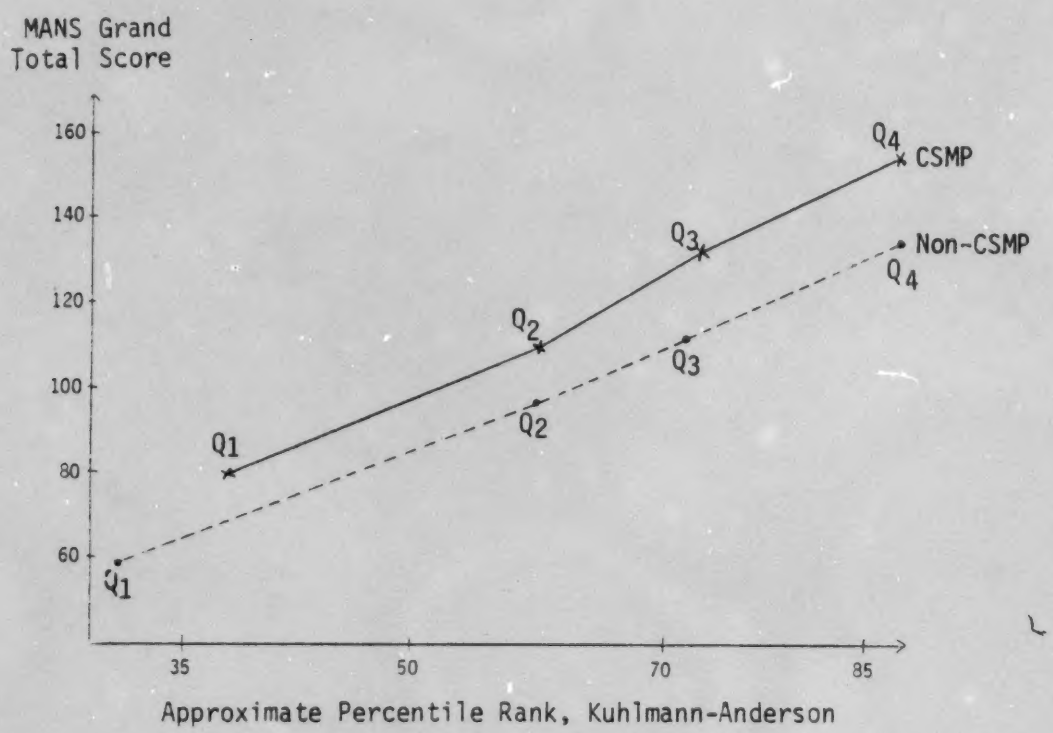


Figure 1 Means by Ability

Figure 1 shows that, as expected, CSMP students did better on this total MANS score, but more importantly that this difference was fairly uniform across ability levels. It is also worth noting that within ability groups, the mean Kuhlmann Anderson scores were nearly equal except that at the lowest ability level the CSMP mean was quite a bit higher than the Non-CSMP mean.

This same procedure was used with each individual scale and the results are shown in the scale-by-scale analysis in Appendix B. Generally the results, like those shown in Figure 1, were consistent across ability levels. But on some scales the differences in favor of CSMP students occurred only at the upper ability levels (for example, Scale #13, Using Hints), while on some other scales the differences occurred mainly at the lower ability levels (for example, Scale 20a), Boxes.

Figure 2 shows graphs for each of the groups of scales discussed previously. Generally there occurs the same pattern of consistent differences across ability levels, except that in the Word Problem scales, and possibly the Number Relationship scales, the differences are larger at higher ability levels. These graphs also corroborate the overall CSMP-Non CSMP differences already reported; virtually no differences on Computation and, at the lower ability levels, on Word Problems; moderate differences in Large Number Computation and Estimation; and large differences in Number Relationships.

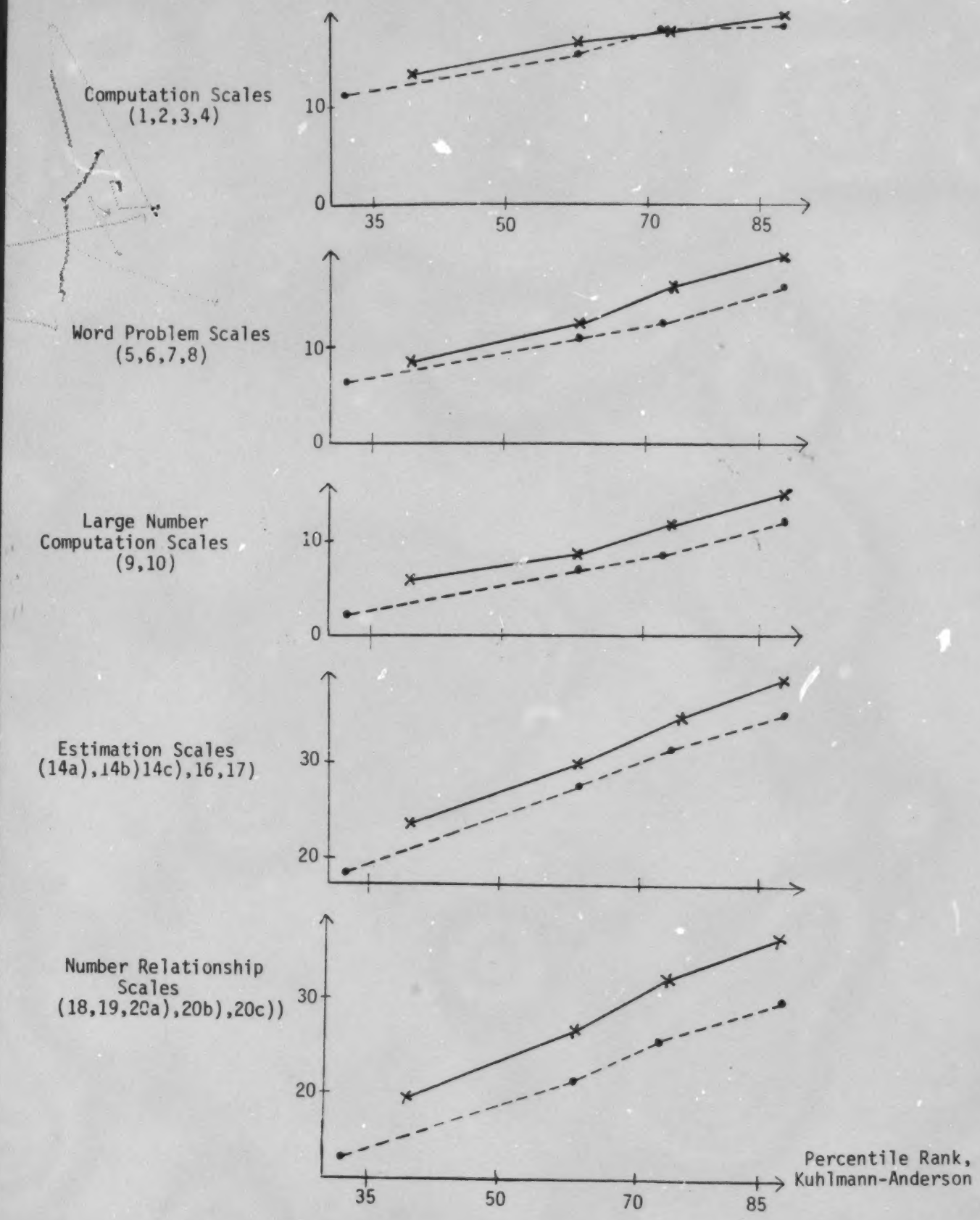


Figure 2 Means by Ability Groups (✕——✕ = CSMP, ●- - - -● = Non-CSMP)

Scores by Sex

For both CSMP and Non-CSMP, separate means were calculated for boys and girls on each of the groups of scales. This information is summarized in Table 8 and, as expected, the results were consistent for boys and for girls; where CSMP boys did better than Non-CSMP boys it was also true that CSMP girls did better than Non-CSMP girls, and usually by about the same amount. On the total of all the MANS, CSMP boys did about 20% better and CSMP girls 18% better than their Non-CSMP counterparts.

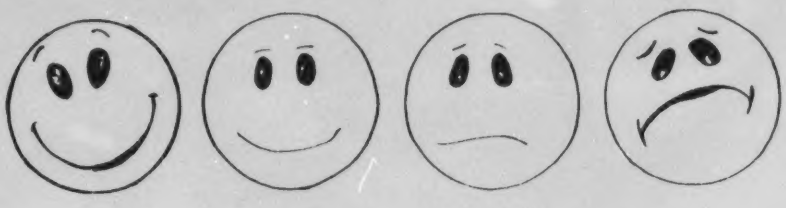
What is interesting about Table 8 is that if one combines CSMP and Non-CSMP boys and then CSMP and Non-CSMP girls, one gets rather interesting sex differences, and this is true regardless of curriculum (this is the column with the heading "Combined" and is based on about 150-160 each of boys and girls). There was virtually no difference between the boys and the girls on the Kuhlmann Anderson and on the Computation and Number Relationship scales. However there were fairly large differences in favor of the boys on the Word Problems and Estimation scales. For 5 of the 25 MANS scales there was a difference in mean scores of between .3 and .4 standard deviations in favor of the boys. These 5 scales were either Word Problem scales (#7, Hits and Misses and #8, Miscellaneous) or Estimation scales (#14a) Estimation-Addition; #16, 2 or 5 or 10; and #17, Which is Larger).

Table 8
Mean MANS Scores by Sex

	Sex	CSMP	Non-CSMP	Combined
Computation	Boys	17.1	15.6	16.3
	Girls	16.7	16.0	16.4
Word Problems	Boys	14.9	12.7	13.8
	Girls	12.8	11.1	11.9
Large Number Computation	Boys	10.5	8.3	9.4
	Girls	9.9	6.6	8.3
Estimation	Boys	33.3	28.9	31.1
	Girls	29.4	26.7	28.2
Number Relationships	Boys	29.7	22.4	26.1
	Girls	27.8	22.1	25.0
Grand Total MANS	Boys	124.5	103.5	114.0
	Girls	113.7	96.4	105.2
Kuhlmann Anderson	Boys	79.2	77.5	78.5
	Girls	79.3	76.6	77.9

Attitude Towards MANS

After each administration of the MANS, students went back through the test to indicate their attitude towards each scale. They did this by marking one of four faces which were shown at the bottom of each test page ("Mark the face that best shows how you felt about this page"). The four faces looked like the following:



These faces were scored 4, 3, 2 and 1 respectively. For each scale a mean "attitude" score was calculated for CSMP and for Non-CSMP students. The total attitude" score across all scales was 68.5 for CSMP versus 66.7 for Non-CSMP students or a mean score per scale of 3.0 for CSMP and 2.9 for Non-CSMP. In other words, the mean score corresponded to the second face from the left.

While CSMP and Non-CSMP attitudinal differences were small overall, there were some larger differences on individual scales. These differences, when they are ≥ 0.3, are shown in Table 9.

Table 9
Scales Producing Different Attitude Responses

Scale	Mean Attitude Score	
	CSMP	Non-CSMP
#3 Computation-Multiplication	3.3	3.6
#4 Computation-Division	2.5	2.9
#14c) Estimation-Division	2.6	2.3
#18 Using Number Machines	3.0	2.5
#19 Solving Number Machines	3.3	2.9
#11 Fractions	2.9	2.5
#9 Large Number Computations (+,÷)	2.8	2.4
#10 " " " (−,×)	3.0	2.7

The scales liked best by CSMP students (the last 6 scales of Table 9) were also scales in which CSMP classes had significantly higher scores. This is not particularly surprising; students tended to most like those scales on which they did best (the correlation between mean scale scores and mean attitude scores was .72). By far the best liked scales for both groups were the familiar and easy ones: Computation-Addition, Subtraction and Multiplication (Scales 1-3) and One-Step Word Problems (Scale 5). The least liked scales involved division (Scales 4, 9, 11, 14c)) or unfamiliar word problems (Scales 7 and 8).

However, the above data does not mean that students who did well on the MANS overall generally tended to have higher attitude scores overall. There was only a small correlation across students between total MANS and total attitude scores (the median correlation for the various test forms being.25) and virtually no relationship between total MANS class scores and total attitude class scores, as Figure 3, below shows. Thus it seems overall attitude is not much affected by overall success, but relative ranking of individual scales very definitely is.

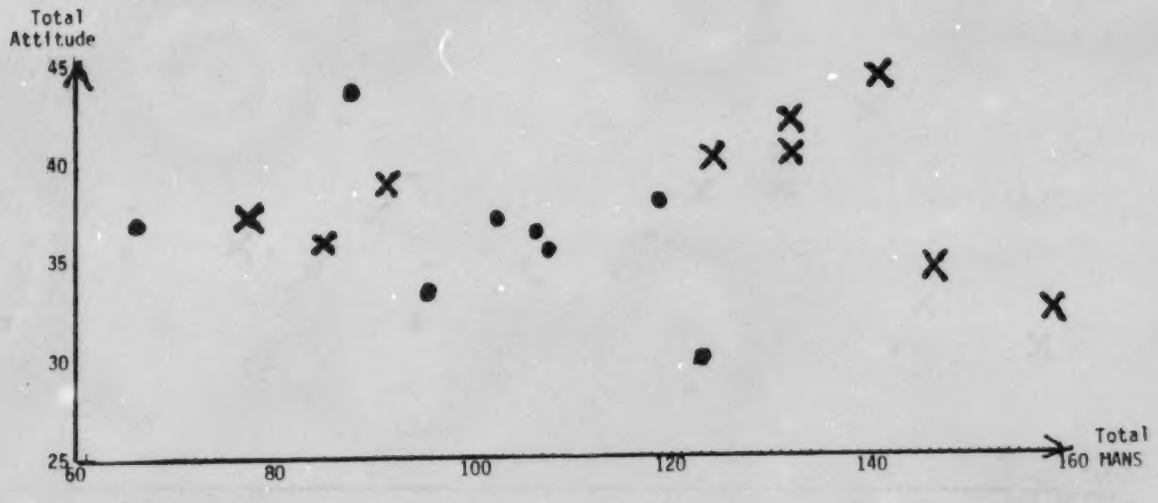


Figure 3 Graph of Class Means: Attitude Toward MANS versus MANS Score
(× = CSMP, • = Non-CSMP)

Key Test Items

It is interesting to note which test items produced the largest differences in percent correct between CSMP and Non-CSMP students. If one takes into account the slightly higher ability scores attained by CSMP students on the Kuhlmann Anderson test (by making a generous rule of thumb adjustment of 5% in favor of Non-CSMP students on each item), there were 7 items on which Non-CSMP students did clearly better. These items are shown below; on all these items Non-CSMP students did at least 13 adjusted percentage points better than CSMP students. The largest difference was 18 points.

$30 \div 6$

$6\overline{)94}$

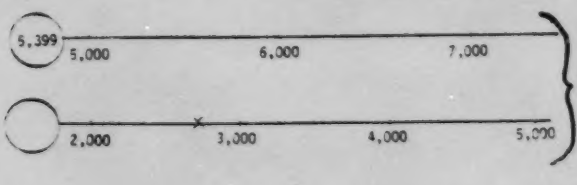
} Scale 4, Computation Division

$\frac{3}{5} - \frac{1}{5}$ Scale 11, Fractions

Which is larger? 38×38 or 39×31

$\frac{2}{3}$ of 500 or $\frac{1}{6}$ of 500

} Scale 17, Which is Larger



} Scale 21, Place Value

There were 39 items on which CSMP students did clearly better than Non-CSMP students (better by at least 13 adjusted percentage points). About 3/4 of these items involved the process of division, either overtly or in a less obvious way (such as in Scale #15, Labelling Number Lines). This is a very high proportion when one considers that division items accounted for only about 40% of all the MANS items. Now many of these items required the student to apply a division or allocation process involving relatively small numbers, but it is still a rather surprising advantage given the result on Scale #1, Computation-Division, in which Non-CSMP students had slightly higher scores.

The items which produced the largest differences (between 25 and 39 adjusted percentage points) are the following:

Scale #6: Joel has 20 sunflower seeds for his pet birds.
Each bird gets the same number of seeds; no seeds are left over.
How many pet birds could Joel have? _____
(List all the possible answers.)

Scale #9: $500 \div 2 = \square$

$989 + \square = 1{,}000$

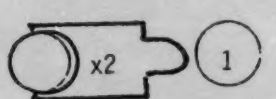
Scale #10: $\square \times 200 = 500$

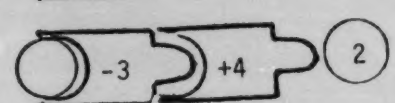
Scale #11: $\frac{1}{2} \times 230 = \square$

$\frac{1}{3}$ of $\square = 12$

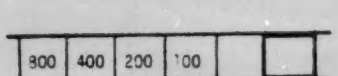
Scale #15:

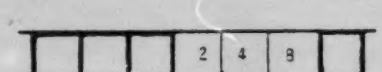
Scale #19:





Scale #20c):





Correlational Data

Three types of correlational data will be noted briefly. First, the KR20 Reliability coefficients were calculated for each scale and are reported in Appendix B. This index is a measure of the internal consistency of a scale, i.e., the degree to which the items are testing basically one underlying ability. For all but 2 of the MANS scales this coefficient was above .66, and the median value was .76 which is fairly high considering the relative brevity of most scales, which contained an average of only about 7½ items. The exceptions were Scale #1, Computation-Addition, whose KR20 may have been low because the scale was so easy for most students that there was little variation in scores, and Scale #14c), Estimation-Division, which was quite difficult and probably elicited a lot of guessing.

Second, correlations between each scale and the Kuhlmann Anderson Test were calculated and these too are reported in Appendix B. The two scales noted above had only moderate correlations of .31 and .40, consistent with the low reliability coefficients. Otherwise there was remarkably little variation, the correlation coefficients falling within the range of .46 to .65. Again, these are fairly substantial correlations given the small number of items in most scales. Another external criterion with which scores could be correlated was the teacher's estimate of each student's "mathematical problem solving ability" (rated from 1 to 5). These correlations also were consistent, between .4 and .6, except as usual for a low .2 for Scale #1 and a surprisingly high .72 for Scale #10, Large Number Computation (÷, x).

The third type of correlations had to do with intercorrelations between MANS tests. Suffice it to say that except for Scale #1, most of the 300 intercorrelations were between .3 and .7 The correlations between MANS X total and MANS Y total were all above .90 for the various forms of each.

Comparisons with Third Grade MANS

Twelve of the 25 MANS scales were similar to scales administered to third grade classes last year in three comparative studies each involving 6 to 12 CSMP and Non-CSMP classes. Evaluation Report 4-B-2 describes the results of that testing; one of the analyses performed was a scale-by-scale analysis of how many of the three sites produced "large" differences, where "large" was a function of the difference in class means divided by the standard deviation (see page 11, Report 4-B-2).

Table 10
Comparison of Third and Fourth Grade Results

	Number of Scale, 3rd Grade	Significant (.05) in Fourth Grade Study	"Large" Effects in at least 2 of the 3 Third Grade Studies
#6 Two Step Word Problems	A4	✓	✓
#9 Large Number Computation (+,-)	A5	✓	✓
#10 Large Number Computation (-,x)			
#11 Fractions	B7	✓	✓
#14b) Estimation-Multiplication	A2	✓	✓
#17 Circle the Larger	B4	✓	✓
#18 Solving Number Machines	A3	✓	✓
#19 Using Number Machines	B2	✓	✓
#15 Labelling Number Lines	A6	✓	
#14a) Estimation-Addition	A2		
#13 Using Hints	A7		
#16 Multiplying by 2, 5 or 10	B1		

The similarity of results is striking; scales which produced consistently large differences in third grade produced significant differences in the fourth grade, and those that didn't in third grade also didn't in fourth grade. In only one of the 12 cases (Scale #15) was this pattern not followed

Another way of making comparisons is by looking at item analysis data. There were a total of 58 items common to the third and fourth grade MANS scales. The mean percent correct on these items in fourth grade was 72 for CSMP students and 64 for Non-CSMP students (after making a small adjustment for differences in the ability levels of the two groups). In third grade the mean percent correct for these same classes of students had been 54 and 46. In other words both groups averaged large gains of almost 20% per item, though CSMP students retained their 8% superiority.

Total MANS Scores

For each class a total MANS score was calculated by adding together all the mean scores of the individual scales. These means were then plotted against the mean Kuhlmann Anderson scores and the graph is shown in Figure 4, below, together with the corresponding regression line. (Appendix B presents similar data for each individual scale.)

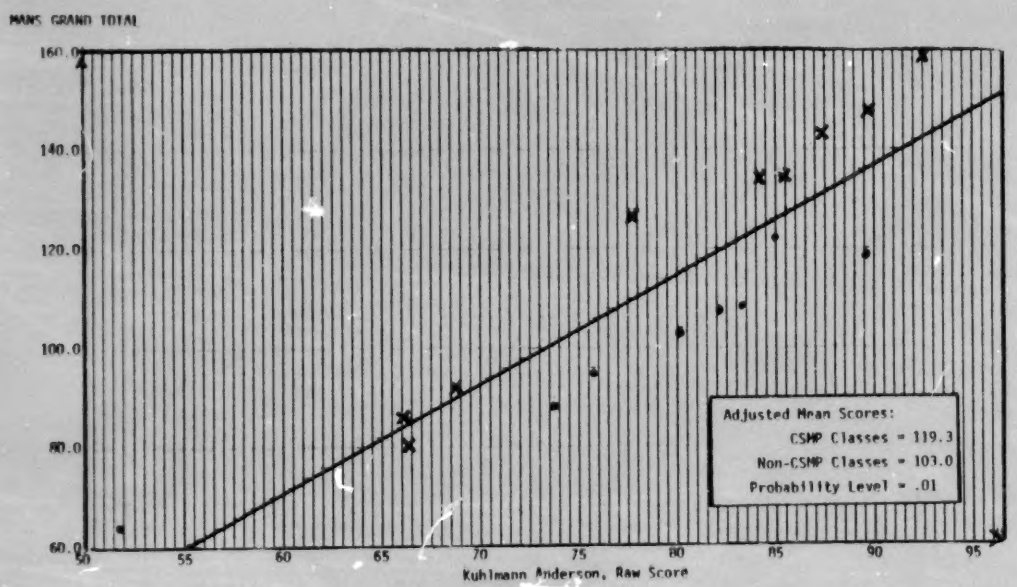


Figure 4 MANS Grand Total Scores

(x = CSMP, • = Non-CSMP)

The significance test run on these class means was really unnecessary; clearly there is a difference in favor of CSMP that is immediately apparent from the graphs. The problem in interpretation is to describe what the total MANS score represents. Obviously it is some kind of measure of mathematical abilities and skills, but whenever a large number of somewhat dissimilar scales like these is combined to form a total score, the labelling of this total score is a problem. If one had combined a

large number of "traditional" scales (like Scales 1-5 and others usually found in standardized tests - common measures, numeration, time, money, geometry, etc.), a very different pattern of class means would probably have resulted. Historically the CSMP-Non-CSMP differences on these scales have been small or non-existent and certainly not as visually apparent as the MANS total score.

What makes the graph so striking is the degree to which these class means fit a linear regression model with the Kuhlmann Anderson. Figure 5 shows the resulting graph of the same class means if one ignores the low ability Non-CSMP class (an outlier with the present data) and draws separate lines. The degree of fit of these lines is astonishing and adds considerable credibility to the CSMP advantage on these MANS Scales.

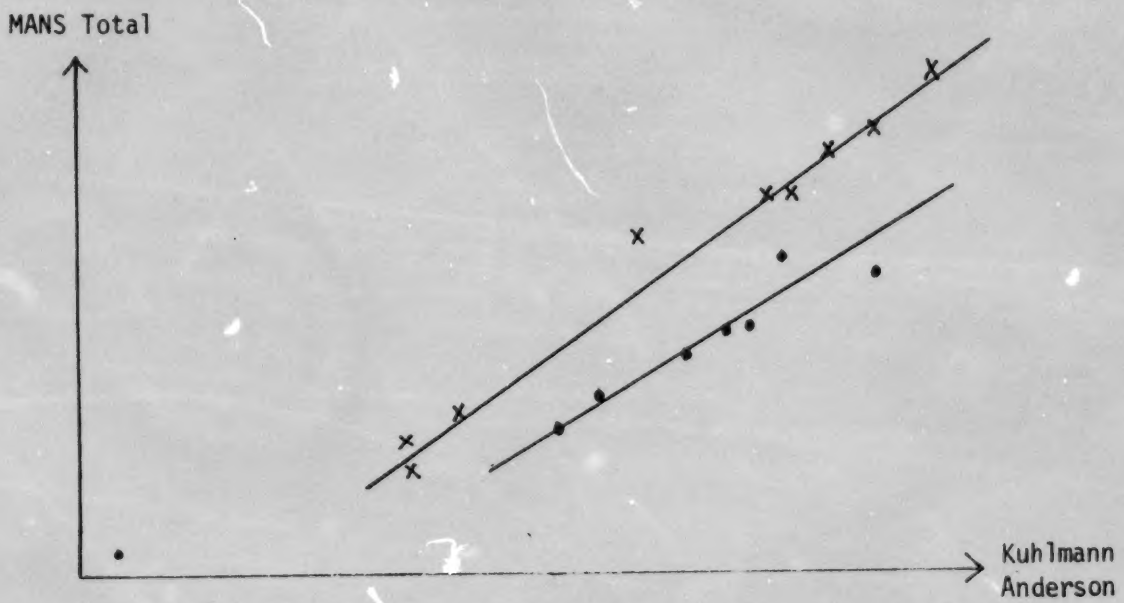


Figure 5 Fourth Grade Class Means
(× = CSMP, • = Non-CSMP)

A casual look at Figure 5 suggests that there may be some place towards the left hand part of the graph at which the regression lines will meet; i.e., at some lower level of ability the CSMP advantage may disappear. Of course this is a completely

unwarranted extrapolation; there are no Non-CSMP classes at the same ability level (Kuhlmann Anderson score) as the three lowest scoring CSMP classes, and at ability levels lower than that there is simply nothing to go on, one way or the other. Thus, whatever conclusions one might draw from the present study should be limited to classes of above average ability (above, say, the 50th percentile).

Summary

In May, 1978, a series of 25 short MANS scales was administered by a specially trained tester, to 9 CSMP and 8 Non-CSMP classes in 3 St. Louis area school districts. A MANS scale typically consisted of about 5-10 open-ended test items dealing with a novel mathematical task that was thought to require thinking processes similar to those underlying the CSMP concepts, (though some scales dealt with rather straightforward computational tasks). A general ability test, the Kuhlmann Anderson Test, was administered to these classes earlier in the year and the results showed the two groups of classes to be of nearly equal ability. The median score for both groups corresponded to about the 70th percentile so that these classes as a whole were above average in ability.

Class mean scores were calculated for each MANS scale, and an Analysis of Covariance was performed using Kuhlmann Anderson scores to adjust for any differences in ability between the two groups of classes. The adjusted mean scores, and p-values from the resulting F test, are shown in Table 10, below. Similar scales are grouped together; actual copies of the scales are in Appendix A and detailed statistical data for each scale is given in Appendix B.

Table 10

Adjusted Mean Scores, MANS Scales

Scales	Number of Items	Adjusted Means CSMP	Adjusted Means Non-CSMP	P-value from F Test (* = Significant at .05)
Computation				
#1 Addition	5	4.4	4.0	.11
#2 Subtraction	5	3.6	3.5	.64
#3 Multiplication	7	5.7	5.3	.09
#4 Division	6	3.3	3.5	.54
Word Problems				
#5 One-Step	8	6.2	5.9	.43
#6 Two-Step	6	4.1	3.1	.01*
#7 Hits and Misses	5	1.8	1.1	.06
#8 Miscellaneous	6	2.7	2.2	.23
Number Relationships				
#18 Solving Number Machines	8	6.1	5.2	.02*
#19 Using Number Machines	9	6.5	4.9	.01*
#20a) Boxes-1	9	6.9	6.0	.01*
#20b) Boxes-2	9	5.9	5.3	.03*
#20c) Boxes-3	7	3.0	1.6	.01*
Estimation				
#14a) Estimation-Addition	8	6.3	5.9	.25
#14b) Estimation-Multiplication	7	5.9	5.3	.01*
#14c) Estimation-Subtraction	7	3.7	3.3	.03*
#16 Multiplying by 2, 5 or 10	13	8.8	8.6	.55
#17 Which is Larger	12	7.5	6.5	.04*
Large Number Compuation				
#9 Large Number Computation (+,-)	9	4.9	3.5	.01*
#10 Large Number Computation (-,x)	8	5.2	4.2	.01*
Other Scales				
#11 Fractions	8	4.3	3.5	.05*
#12 Measuring	6	3.2	2.7	.05*
#13 Using Hints	5	1.9	1.4	.10
#16 Number Line Labelling	6	3.8	2.8	.01*
#21 Place Value	8	5.0	4.9	.65

Table 10 shows that the CSMP classes did significantly better on all the Number Relationship and Large Number Computation scales; on some of the Estimation and "Other" scales; and on one of the more traditional Computation and Word Problem scales. Non-CSMP classes did better on only one scale, Computation-Division, and the difference was not significant.

These results are not surprising. Based on the relative emphasis in CSMP versus traditional curricula on various kinds of problems, one might have predicted this result, i.e., the largest differences to occur in the Number Relationship and Large Number Computation scales and the smallest differences to occur in the Computation and Word Problem scales. This is not meant to minimize the impact of the

results in any way; it is no small achievement to attain significant differences on tasks which are new to students. In fact this predictability, together with the remarkable consistency of the results from year-to-year (see page 27 for third grade versus fourth grade), add much to the credibility of the results; students who study CSMP do seem to be consistently better at certain kinds of tasks than Non-CSMP students, and they are the kind of tasks on which one might have expected such differences.

Other findings were the following:

1. When mean scores were calculated for students at different ability levels, it was found that when CSMP scores were higher overall on a given scale, they were also higher at each ability level. There were occasional exceptions on individual scales, but this was certainly true when scales were combined into groups and when Total MANS scores were computed (see the Figures on pages 18 and 20).

2. CSMP boys outperformed Non-CSMP boys by about the same amount (18-20%) as CSMP girls outperformed Non-CSMP girls. Less expected perhaps was the finding for both CSMP and Non-CSMP students, that while there was virtually no difference between boys and girls on Kuhlmann Anderson scores and on the MANS Computation scales, boys outperformed girls by about one-third of a standard deviation - a significant difference - on the Word Problem and Estimation scales. (See page 21)

3. There was no relationship between total MANS score and total Attitude score (calculated by summing the individual student ratings on each scale); in other words students who did well on the MANS did not rate the test as a whole any higher than students who did poorly. Furthermore, there was no difference overall between CSMP and Non-CSMP on MANS rating. In terms of relative ranking, students

tended to rank highest those scales in which they did best. For example, CSMP students rated the Number Machine Scales (#18,19) and the Large Number Computation (#9,10) much higher than did Non-CSMP students and they also perfromed significantly better on these scales. (See page 22)

4. Out of the 181 MANS test items there were a total of 7 rather unrelated items on which Non-CSMP students did much better than CSMP students, differences ranging from 13 to 18 percentage points (percentages slightly adjusted to accommodate differences on Kuhlmann Anderson scores). There were 39 items on which CSMP students did much better, differences ranging from 13 to 39 points. A disproportionate number of these, about 3/4, dealt in one way or another with the process of division, an interesting finding in view of the fact that the Computation-Division scale was the only one in which Non-CSMP students had a higher score. (See page 24)

5. About half the scales were identical or very similar to scales administered to these and other classes of students last year when they were third graders. The results were strikingly similar; scales which produced large differences in third grade continued to do so in fourth grade and scales which did not produce large differences in third grade continued not to in fourth grade (except Scale #15, Labelling Number Lines, which did). On an item analysis of 58 identical items both CSMP and Non-CSMP students had an average increase of about 20 in percent correct, CSMP students maintaining an 8 point advantage from year to year. (See page 27)

6. The Kuhlmann Anderson test was a fairly good predictor of MANS scores for individual students (correlations usually between .4 and .7) and an excellent predictor of class mean scores (though less so for the Computation scales and some of the Word Problem scales). In fact, when class means for Total MANS were plotted against class means for Kuhlmann Anderson, and separate regression lines plotted for CSMP and for Non-CSMP classes, every class but one was found to lie right on or very close to the appropriate regression line. This seems to indicate that the teacher effect (as opposed to the "CSMP" effect) on total MANS score is small, though of course there were many large differential effects on individual scales. (See page 28)

It is important in this kind of study to ask two questions:

a) Was it really the CSMP curriculum that produced these differences?

b) Should one expect these results to occur elsewhere?

With respect to the first question dealing with the problem of internal validity, a necessary condition for one to be able to give an affirmative answer is that teachers and classes be randomly selected. As is usual in educational evaluation, this condition was not met. Thus one has to fall back on a subjective judgment of whether the CSMP and Non-CSMP classes and teachers were comparable. There seems little doubt that the classes were comparable; they were from nearby schools and had similar ability levels as measured by Kuhlmann Anderson scores (adjustments in the Analysis of Covariance procedure were quite small). As to the comparability of the teachers there is less to go on, and in any case it is not all clear which teacher characteristics lead to improved student performance. The CSMP teachers were each observed several times; as a group they seemed to be a fairly typical group of teachers - some good and some not so good - who generally implemented the program about as intended.

There is one area in which some doubt must be raised. Questionnaire and interview data was collected from all fourth grade CSMP teachers, including those in this study (Evaluation Report 5-C-1), and a subset of those questions was also asked of the Non-CSMP teachers. One of these questions dealt with how much time was taken each day for math class. The CSMP classes in this study averaged about 55 minutes in length versus the Non-CSMP classes' average of about 40 minutes. How much of this was "time on target", and how the extra time was used, is not known. CSMP teachers <u>did</u> report supplementing the program with basic computational practice, which consumed usually 10 to 20 percent of the math time. To what extent should one consider this supplementing as a part of CSMP math for the purposes of question a), above? The situation is complicated by the fact that Non-CSMP teachers reported about the same degree of supplementing for <u>their</u> (textbook) math programs. In any case the differential between CSMP and Non-CSMP math time was not found in comparative studies at earlier grade levels and might even be viewed as a natural and desirable result of CSMP usage.

With respect to the second question, regarding the generalizability of the results, it must be again emphasized that the group of classes involved in this study was well above average in ability. There were a number of low ability students <u>within</u> these classes, and for these students similar comparative results were obtained as for higher ability students. But this does not help answer the question of what one should expect in low ability fourth grade <u>classes</u>. In addition it must be noted that this study involved only 9 CSMP classes in one area. Further studies will be conducted with broader participation, but until this is done, the present study seems a little too narrowly based to yield a very definite affirmative answer to the second question.

Appendix A

Description of MANS Scales

Brief descriptions of each scale and accompanying directions are given on the following pages, along with reduced copies of the actual pages in the student test booklet. The tests were administered by specially trained testers who followed a script considerably longer than is indicated here by the "Abbreviated Directions". A time limit was enforced only for Scales 13, 14, 16, and 17 (the scales on which working out the whole calculation was not supposed to be done). For these scales students were warned after half the time was up. The other scales were essentially untimed; the tester began the next scale when everyone was finished or no further progress was being made.

The table below summarizes the various MANS 1978 scales.

Name of Scale	Number of Items	Number of Minutes: Directions	Number of Minutes: Working	Test Booklet[1]	Description Page[2]	Results Page[3]
#1 Computation: Addition	5	¼	5	Y	A2	B4
#2 " Subtraction	5	¼	5	Y	A3	B6
#3 " Multiplication	7	¼	6	X	A4	B8
#4 " Division	8	¼	6	X	A5	B10
#5 Word Problems: One-Stage	8	¼	6	X	A6	B12
#6 " " Two-Stage	6	½	6	Y	A7	B14
#7 " " Hits and Misses	5	½	6	Y	A8	B16
#8 " " Miscellaneous	6	¼	6	X	A9	B18
#9 Large Number Computation (+,÷)	9	¼	5	Y	A10	B20
#10 " " " (−,×)	8	¼	5	Y	A11	B22
#11 Fractions	8	½	4	Y	A12	B24
#12 Measuring	6	1½	6	X	A13	B26
#13 Using Hints	5	1½	2½	Y	A14	B28
#14a) Estimation: Addition	8	2	1½	Y	A15	B30
#14b) " Multiplication	7	¼	1½	Y	A16	B32
#14c) " Division	7	¼	1½	Y	A16	B34
#15 Labelling Number Lines	6	1½	6	X	A17	B36
#16 Multiplying by 2, 5 or 10	13	1	2½	X	A18	B38
#17 Which is Larger	12	2	2½	X	A19	B40
#18 Solving Number Machines	6	3	6	X	A20	B42
#19 Using Number Machines	9	2½	5	X	A21	B44
#20a) Boxes(1)	9	3½	4	Y	A22	B46
#20b) Boxes(2)	9	½	4	Y	A23	B48
#20c) Boxes(3)	7	1	4	Y	A23	B50
#21 Place Value	8	1	4	Y	A24	B52

[1]Where there are brackets joining two entries, it means half the test booklets contained one of the two scales, the other half of the booklets contained the other scale. Thus any time a scale is in a bracket, it means that that scale was administered to a random half of the students in each class.

[2]Page number in this Appendix A

[3]Page number in Appendix B

SCALE #1: Computation - Addition

NOTE: In each class, half the students did this scale, the other half did Scale #2, Computation-Subtraction.

WORKING TIME: Untimed, but usually required about 5 minutes

Add

$$\begin{array}{r} 347 \\ +128 \\ \hline \end{array}$$

$793 + 8 =$ ☐

$$\begin{array}{r} 708 \\ +293 \\ \hline \end{array}$$

$574 + 258 =$ ☐

$$\begin{array}{r} \$43.57 \\ 6.89 \\ .62 \\ 7.15 \\ \hline \end{array}$$

SCALE #2: Computation - Subtraction

NOTE: In each class, half the students did this scale, the other half did Scale #1, Computation - Addition.

WORKING TIME: Untimed, but usually about 5 minutes.

Subtract

606
-482

203 - 8 = ☐

738
-169

680 - 155 = ☐

1024
- 727

SCALE #3: Computation - Multiplication

NOTE: In each class, half the students did this scale, the other half did Scale #4, Computation-Division

WORKING TIME: Untimed, but usually required about 6 minutes

Multiply

5 x 8 = ☐ 3 x 35 = ☐ 4 x 210 = ☐

9
x7

43
x7

623
x4

314
x34

SCALE #4: Computation - Division

NOTE: In each class, half the students did this scale, the other half did Scale #3, Computation-Multiplication.

WORKING TIME: Untimed, but usually required about 6 minutes

DIVIDE

$64 \div 2$ means the same as $2\overline{)64}$

$30 \div 6 =$ ☐ $\qquad 7\overline{)56}$

$219 \div 3 =$ ☐ $\qquad 4\overline{)128}$

$144 \div 6 =$ ☐ $\qquad 6\overline{)94}$

SCALE #5: Word Problems - One Step

NOTE: In each class, half the students did this scale, the other half did Scale #8, Word Problems-Miscellaneous.

ABBREVIATED DIRECTIONS:

You read each problem yourself, figure out the answer and put it in the blank.

If there's a word you can't read, just hold up your hand and I'll read it to you.

WORKING TIME: Untimed, but usually required about 6 minutes

1. Bill started with 22 candy bars.

   He sold 17 of them.

   How many candy bars did he have then? _____

2. Andy bought 4 cartons of cola.

   Each carton holds 8 bottles.

   How many bottles did he buy altogether? _____

3. Jane has 25¢.

   Pencils cost 7¢ each.

   How many pencils can Jane buy? _____

   How much money will she have left over? _____

4. Ann has 31¢.

   She wants to buy a game that costs 45¢.

   How much more money does she need? _____

5. Joe started with 40 marbles.

   His father gave him 8 more.

   Then his mother gave him 9 more.

   How many marbles did he have then? _____

6. There are 4 men who take turns parking cars.

   Altogether they will park 24 cars.

   How many will each man park? _____

7. Tom has 3¢ more than Ann.

   Tom has 5¢ less than John

   If John has 20¢, how much does Ann have? _____

SCALE #6: Word Problems – Two Step

NOTE: In each class, half the students did this scale, the other half did Scale #7, Word Problems (Hits and Misses)

ABBREVIATED DIRECTIONS:

You read each problem yourself, figure out the answer and put it on the blank.

If there's a word you can't read, just hold up your and and I'll read it to you.

The answer blank beside the very last question is very long. That's because you are supposed to write down all the possible answers to that problem.

WORKING TIME: Untimed, but usually required about 6 minutes

1. On Saturday Amy and Susan made $13 selling lemonade.
   On Sunday they made $5.
   They put their money together and divided it evenly.
   How much did each girl get? _____

2. Peter bought 4 blue pencils and 3 red pencils.
   Each pencil costs 10¢.
   What was the total cost? _____

3. Pam's allowance is $1.00 each week.
   She always spends 20¢ and saves the rest.
   How much will she save in 5 weeks? _____

4. There are 40 apples in my barrel now.
   I will eat 2 apples every day.
   How many apples will be left in my barrel after 5 days? _____

5. Jim has $10 in his bank now.
   Each week he will add $5 to his bank.
   In how many weeks will he have $30 in his bank? _____

6. Joel has 20 sunflower seeds for his pet birds.
   Each bird gets the same number of seeds; no seeds are left over.
   How many pet birds could Joel have? _____
   (List all the possible answers.

SCALE #7: Word Problems—Hits and Misses

NOTE: In each class, half the students did this scale, the other half did Scale #6, Word Problems (Two Step)

ABBREVIATED DIRECTIONS:

You read each problem yourself, figure out the answer and put it on the blank.

If there's a word you can't read, just hold up your hand and I'll read it to you.

The answer blank beside the very last question is very long. That's because you are supposed to write down all the possible answers to that problem.

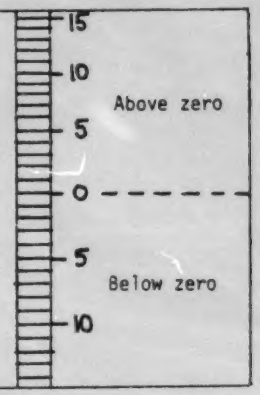
WORKING TIME: Untimed, but usually required about 6 minutes

Read about the Game in Box A. Then do the problems in Box B.

A

In this Game we throw bails at a target.

The scoreboard looks like this →



If we hit the target, we go up 5 points:

HIT: UP 5 POINTS

If we miss the target we go down 2 points:

MISS: DOWN 2 POINTS

B

1. Bill started with a score of 0.
   Then he had 2 hits and 3 misses.
   What was his score then? _____

2. Mary started with a score of 7 below zero.
   Her next 2 throws were hits.
   What was her score then? _____

3. John started with a score of 0.
   Then he had 1 hit and 4 misses.
   What was his score then? _____

4. Toni started with a score of 9 below zero.
   Her next 4 throws were misses.
   What was her score then? _____

5. Hank started with a score of 0.
   Then he had 3 throws.
   What could his score be then? _____
   (List all the possible answers.)

SCALE #8: Word Problems - Miscellaneous

NOTE: In each class, half the students did this scale, the other half did Scale #5, Word Problems-One Step

ABBREVIATED DIRECTIONS:

You read each problem yourself, figure out the answer and put it in the blank.

If there's a word you can't read, just hold up your hand and I'll read it to you.

WORKING TIME: Untimed, but usually required about 6 minutes

1. Mr. Jones owns a truck company.
   His trucks can carry 3 cars each.
   He is loading 14 cars onto trucks.
   How many trucks will he need? _____

2. It takes 4 men to keep an oil well working.
   There are 18 men ready to work.
   How many oil wells can they keep working? _____

3. We buy 3 pounds of nuts for 25¢.
   How many pounds can we buy for $1.00? _____

4. There are 70 ounces in 6 cans of cola.
   How many ounces are there in 3 cans? _____

5. Amy sold 10 tickets to her school dance.
   Altogether she got $3.50 for these tickets.
   How much did each ticket cost? _____

6. The meter on Bill's car shows $247\frac{9}{10}$ miles. | 2 | 4 | 7 | 9 |
   He drives another $\frac{1}{10}$ of a mile.
   What will the meter show then? | | | | |

SCALE #9 - Large Number Computation (+, ÷)

NOTE: In each class, half the students did this scale, the other half did Scale #10, Large Number Computation (−, ×).

WORKING TIME: Untimed, but usually required about 5 minutes

Add

375 + ☐ = 600

☐ + 125 = 250

989 + ☐ = 1,000

Divide

500 ÷ 2 = ☐

700 ÷ 10 = ☐

300 ÷ 50 = ☐

800 ÷ ☐ = 200

☐ ÷ 3 = 120

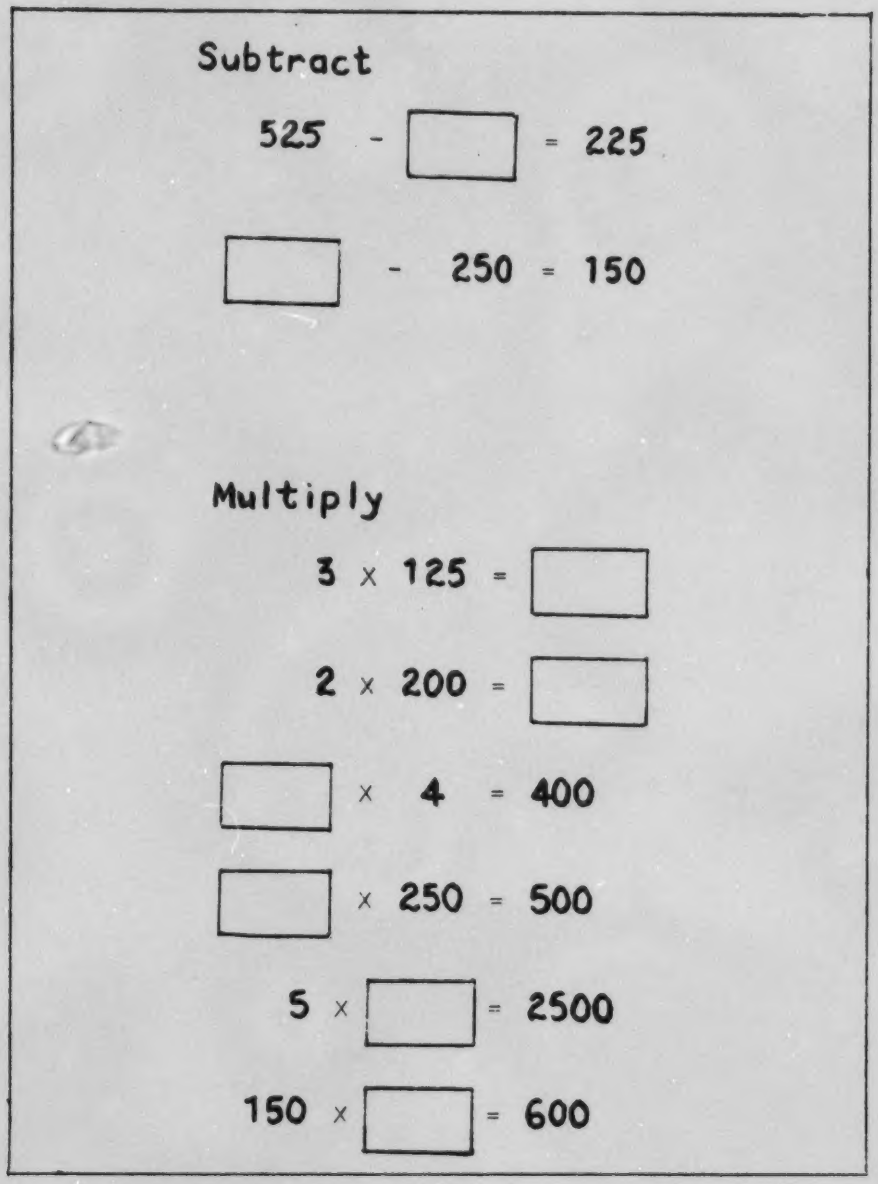
600 ÷ ☐ = 4

5

SCALE #10: Large Number Computation (−, ×)

NOTE: In each class, half the students did this scale, the other half did Scale #9, Large Number Computation (+, ÷)

WORKING TIME: Untimed, but usually required about 5 minutes



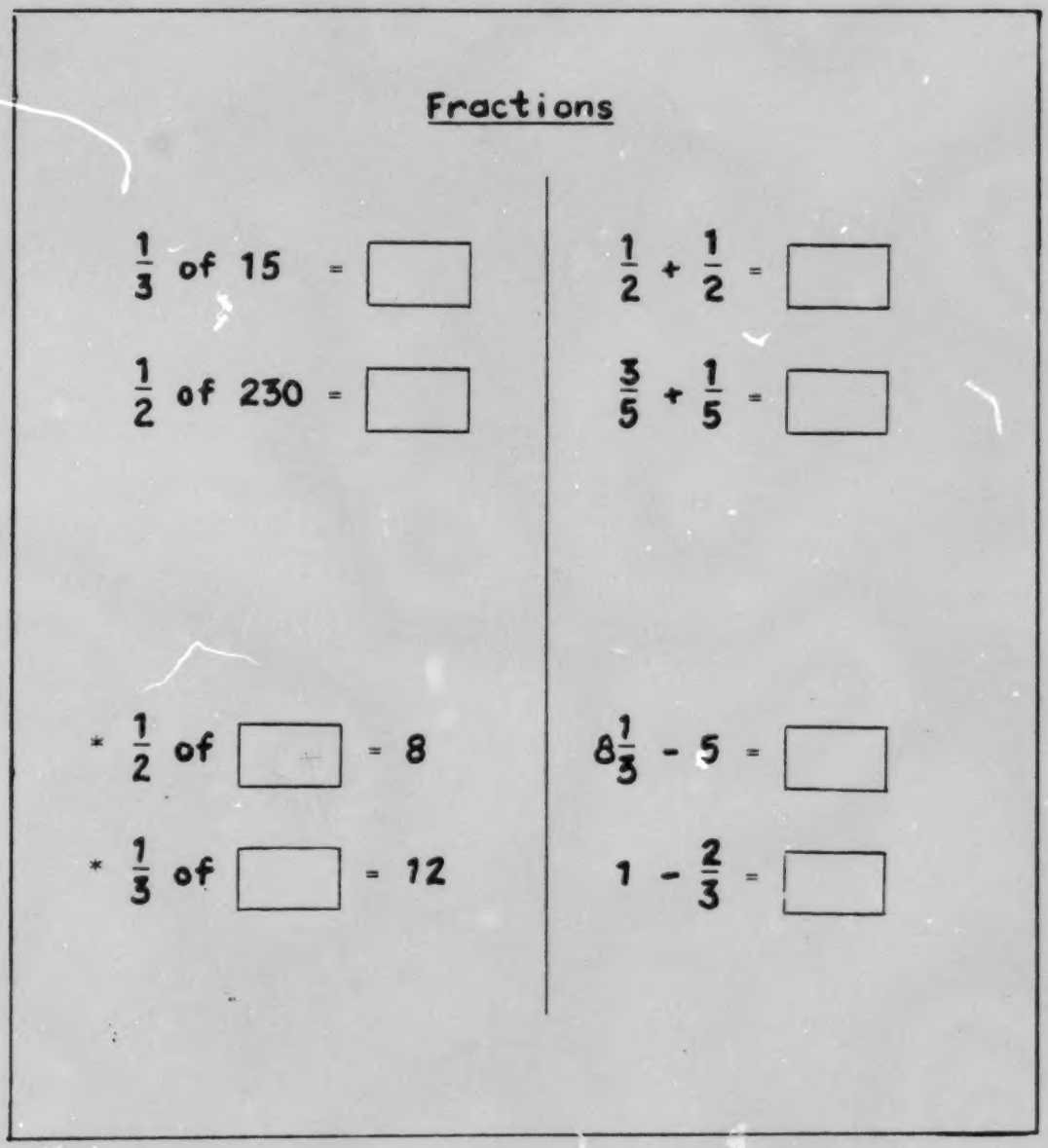
SCALE #11: Fractions

ABBREVIATED DIRECTIONS:

Be careful with the two problems with the stars beside them.

For these two questions the answers are <u>given</u> and we have to figure out what number we <u>started</u> with.

WORKING TIME: Untimed, but usually required about 4 minutes.



5.1

SCALE #12 - Measuring

ABBREVIATED DIRECTIONS:

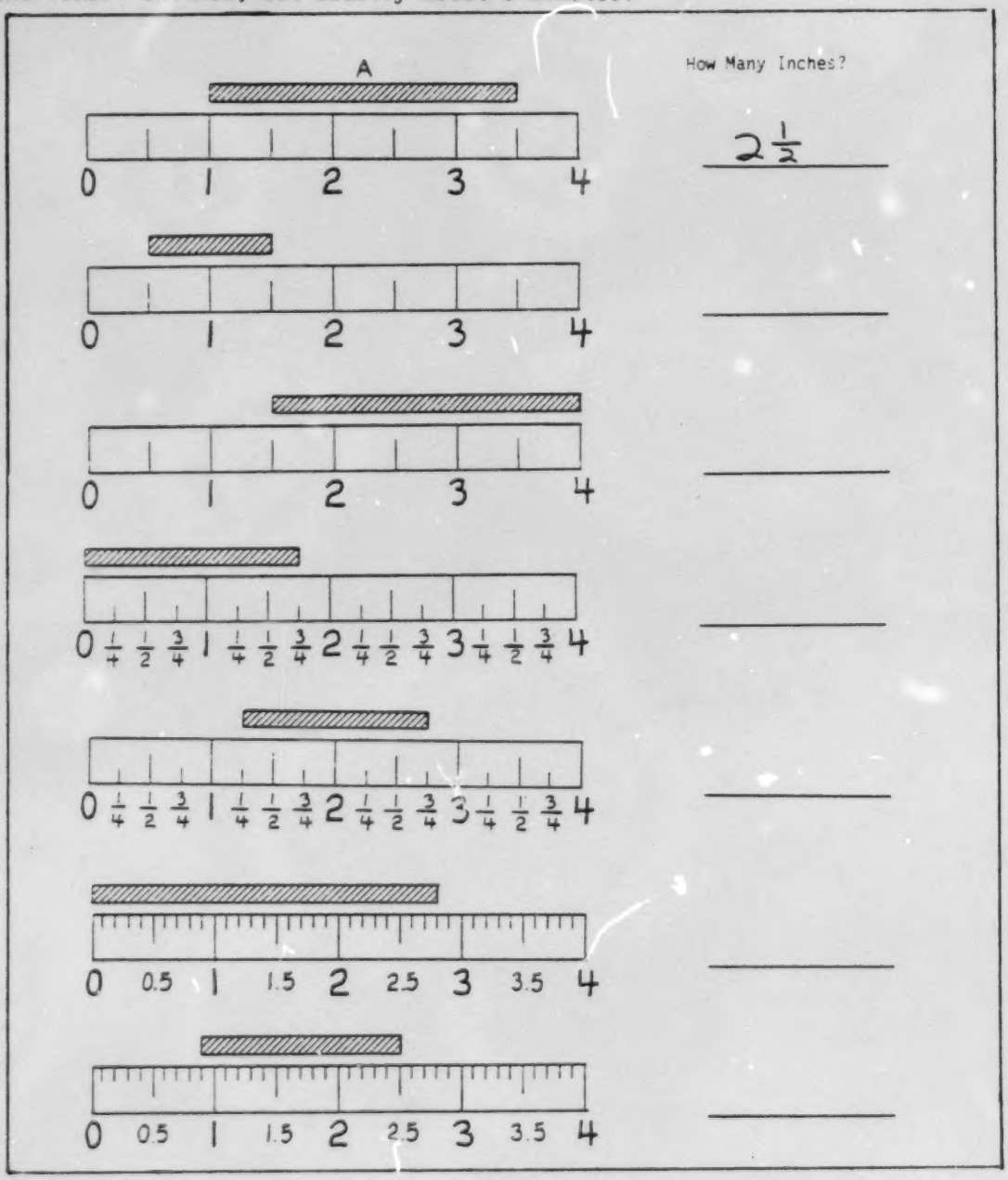
We have to figure out how many inches long each of these rods is.

Rod "A" starts at the 1 inch mark and goes to the mark halfway between the 3 and 4.

It takes 2 inches to get from the 1 mark to the 3 mark. Then there's another ½ inch to get to the mark halfway between 3 and 4. So that's 2½ inches altogether.

All the rulers are 4 inches long, but some are marked differently. The two rulers at the bottom of the page have marks for every tenth of an inch.

WORKING TIME: Untimed, but usually about 6 minutes.



SCALE #13 - Using Hints

ABBREVIATED DIRECTIONS:

On this page there are some problems that a student did.

The student's answer to the first problem was 6,476.

This answer was right but then he found out that he had copied the first number wrong: 637 should have been 647.

Now he has to correct his answer. 647 is 10 more than 637 so the new answer to the problem must also be 10 more. So instead of 6,476 it will be 6,486.

Don't do each calculation all over again. Look at the error the student made in copying and then change the answer to exactly make up for his copying error.

You won't have much time, don't do the whole calculation all over again. Just correct the answer the student got.

WORKING TIME: Exactly 2½ minutes

Add

637 293 876 1,428 135 689 332 771 203 +1,112 6,476	6,577 259 8,314 423 975 547 311 121 463 + 337 16,327
72 67 78 91 56 32 10 13 22 84 + 79 574	958 813 294 875 924 199 267 138 261 422 + 198 5,346

Multiply

6 x 74 = 444

20 x 16 = 320

SCALE #14: Estimating Intervals

NOTE: This scale has four pages. The first page consists of only the two sample items and is shown in the box beside ABBREVIATED DIRECTIONS, below. The second page (Addition) is then shown below and the third and fourth pages (Multiplication and Division) are shown on the next page. Each of three test pages was separately timed.

ABBREVIATED DIRECTIONS:

Look at the first practice problem:

We don't know the exact answer, but we know it must be larger than 500, and it must be samller than 1000, so the "x" went between the 500 and 1000.

Now look at the next calculation. It says 48 + 49. Where do you think we would mark our "x"?

We know it's larger than 50. But 48 + 49 is smaller than 50 + 50 so our answer must be smaller than 100.

So you put an "x" there, between 50 and 100.

For each problem, you should just look at it and then decide quickly where the "x" should go. Don't calculate the exact answer - that would take too long and you wouldn't have time to finish.

Practice

483 - 357	0	10	50	100	500	X	1000
48 + 49	0	10	50	100	500		1000

WORKING TIME: Exactly 1½ minutes for each of the three test pages.

ADDITION

185 + 97	0	10	50	100	500	1000
24 + 24	0	10	50	100	500	1000
59 + 39	0	10	50	100	500	1000
479 + 86	0	10	50	100	500	1000
279 + 165	0	10	50	100	500	1000
19 + 29	0	10	50	100	500	1000
257 + 294	0	10	50	100	500	1000
19 + 19 + 19	0	10	50	100	500	1000

SCALE #14: Estimating Intervals (con't)

MULTIPLICATION

5 x 109	0	10	50	100	500	1000
8 x 19	0	10	50	100	500	1000
40 x 10	0	10	50	100	500	1000
4 x 23	0	10	50	100	500	1000
11 x 50	0	10	50	100	500	1000
2 x 49	0	10	50	100	500	1000
4 x 29	0	10	50	100	500	1000

DIVISION

300 ÷ 4	0	1	10	20	100
200 ÷ 9	0	1	10	20	100
190 ÷ 10	0	1	10	20	100
1 ÷ 2	0	1	10	20	100
150 ÷ 13	0	1	10	20	100
150 ÷ 100	0	1	10	20	100
180 ÷ 21	0	1	10	20	100

SCALE # 15: Labelling Number Lines

ABBREVIATED DIRECTIONS:

Look at the number line at the top of the page.

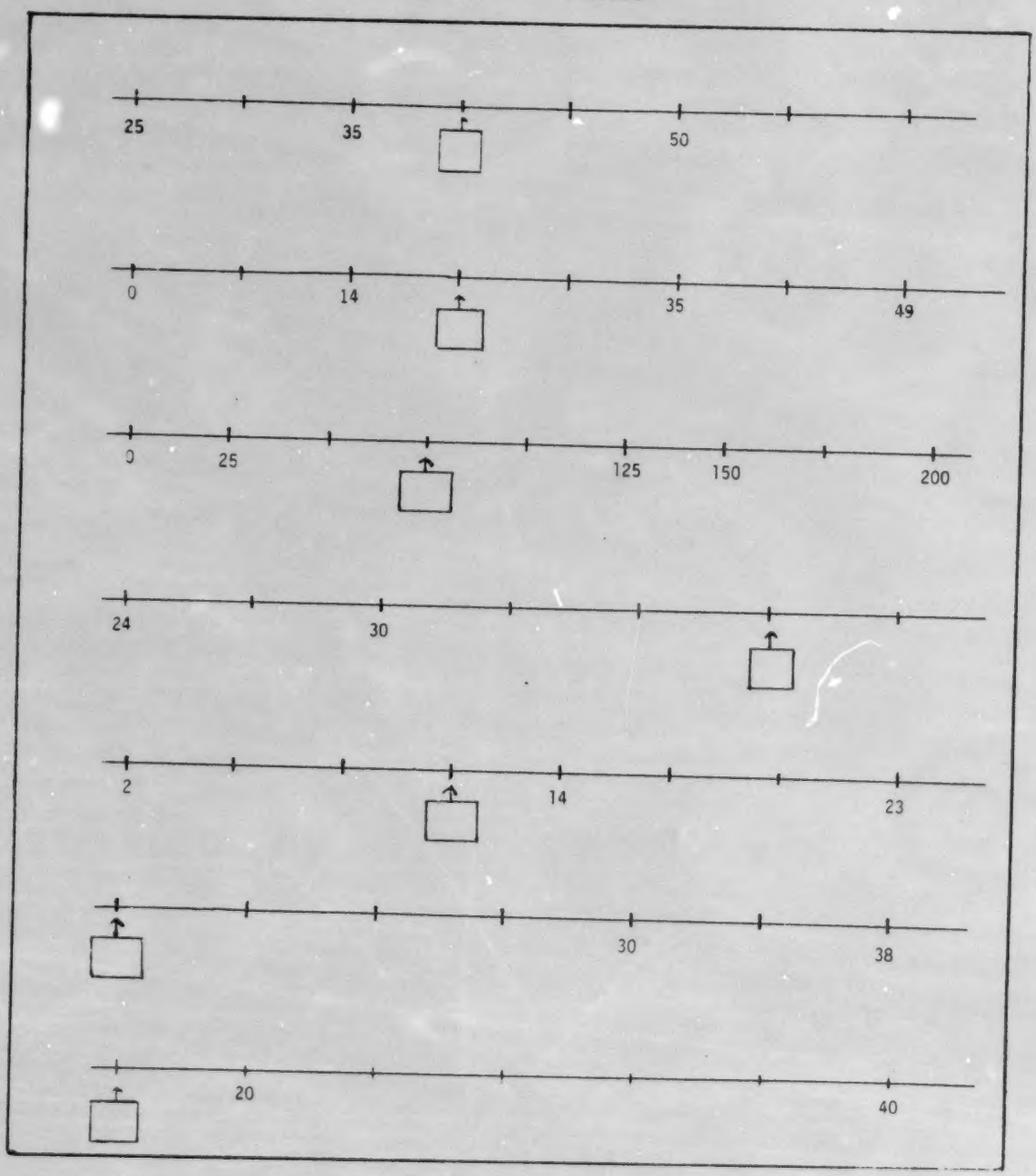
What will the mark between the 25 and the 35 be for?

So how much are the marks going up by each time?

So what will we put in the box after 35? Right 40; 35 + 5 = 40.

For each problem, you <u>first</u> have to figure out how much the marks are going by. <u>Then</u> you decide what goes in the box.

WORKING TIME: Untimed, but usually about 6 minutes.



SCALE #15: Multiplying by 2, 5 or 10

ABBREVIATED DIRECTIONS:

Let's look at the first one.

You decide if 60 is <u>about</u> 2x as large, 5x as large or 10x as large as 31.

None of them is <u>exactly</u> right, but one is a much better answer than the others.

What is the best answer? 2 is, because 60 is <u>about</u> 2 times as large as 31.

In each of the questions write 2 or 5 or 10 in the blank, whichever makes the most sense.

Don't try to calculate the exact answers because then you won't have time to do all the problems.

WORKING TIME: Exactly 2½ minutes

2 or 5 or 10

60 is about _____ times as large as 31

51 is about _____ times as large as 5

16 is about _____ times as large as 7

65 is about _____ times as large as 12

200 is about _____ times as large as 21

100 is about _____ times as large as 19

98 is about _____ times as large as 50

45 is about _____ times as large as 8

602 is about _____ times as large as 298

503 is about _____ times as large as 49

59 is about _____ times as large as 31

12 is about _____ times as large as 2½

20½ is about _____ times as large as 10½

5½ is about _____ times as large as ½

SCALE #17: Which is Larger

ABBREVIATED DIRECTIONS:

Look at the two computation questions at the top of the page.

Look at the first number of each computation, 12 and 6.

Now look at the second number - 510 is a little bigger than 500 - but it certainly isn't twice as big.

So the first computation - 12 x 500 is larger. We know that it's larger without working out the exact answers. So we draw a circle around 12 x 500 to show that it's larger.

For each problem you look at the two computation questions and decide which answer will be larger. Don't calculate the exact answer each time; that would take too long and you wouldn't have time to finish all the problems.

WORKING TIME: Exactly 2½ minutes.

12×500	6×510
$899 + 899 + 899 + 899$	899×5
$8{,}000$	$8 \times 16{,}000$
38×38	39×31
3×162	4×160
$6{,}284 \div 7$	$6{,}284 \div 8$
$599 \div 2$	300
$800 \div 3$	$900 \div 31$
$25{,}000 \div 60$	$26{,}000 \div 90$
$\frac{3}{8}$	$\frac{3}{4}$
$\frac{1}{2}$	$\frac{2}{5}$
$\frac{1}{3}$ of 19,000	$\frac{1}{2}$ of 19,000
$\frac{2}{3}$ of 500	$\frac{1}{5}$ of 500

SCALE #18: Solving Number Machines

ABBREVIATED DIRECTIONS:

Look at the number machine in Box A.

When you put 3 into the machine, the result is 6. When you put 2 into the machine the result is 4. When you put in 4, the result is 8.

What <u>is</u> the machine in Box A doing to each number? (Multiplying by 2)

Now the next number to be put in is 5. So what is m multiplied by 2? (10)

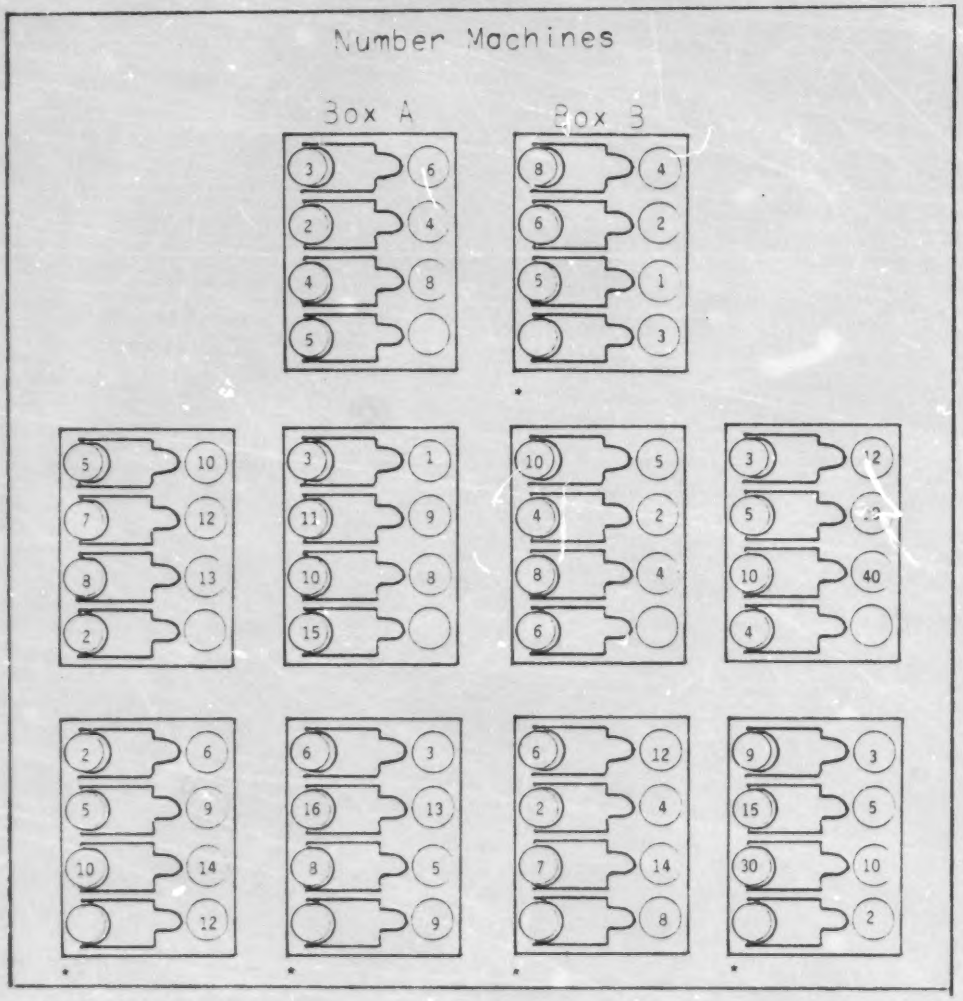
Now look at Box B. It has a different machine.

What is it that this machine Box <u>B</u> is doing each time? (Subtracting 4)

This time we know the <u>result</u> was 3, and we have to figure out what number we started with. What number must we have started with so that when we subtract 4 we end up with 3? (7; 7 subtract 4 is 3)

Be careful with the rest of these. Each box is for a <u>different</u> machine - but it'll always do the same thing to all the numbers inside the box. It might be adding or subtracting, or multiplying or dividing. Watch out for the four problems in the bottom row. They all have stars below them; that means you have to figure out what number we <u>started</u> with.

WORKING TIME: Untimed, but usually required about 6 minutes.



SCALE #19 - Using Number Machines

NOTE: This scale has two pages. The first page consists of only the three samples and is shown in the box beside the ABBREVIATED DIRECTIONS. The other page had the test items and is shown below.

ABBREVIATED DIRECTIONS:

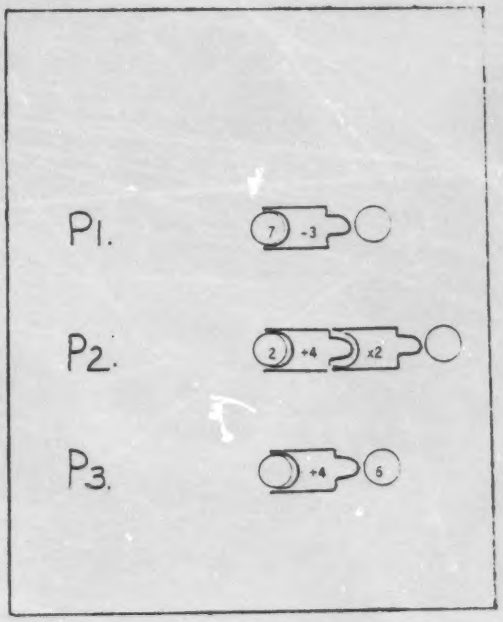
Look at problem P1. The "-3" tells us that when we put a number into the machine it automatically subtracts 3 and then shows the result on the other side.

What will be the result for P1? Right 4; 7 goes in, we subtract 3 from 7 and that gives a result of 4.

Look at P2. We put 2 in the first machine. It adds 4 giving a result of 6 and so 6 goes into the second machine. The second machine multiplies this 6 by 2, so the final result is 12. You put 12 in the circle.

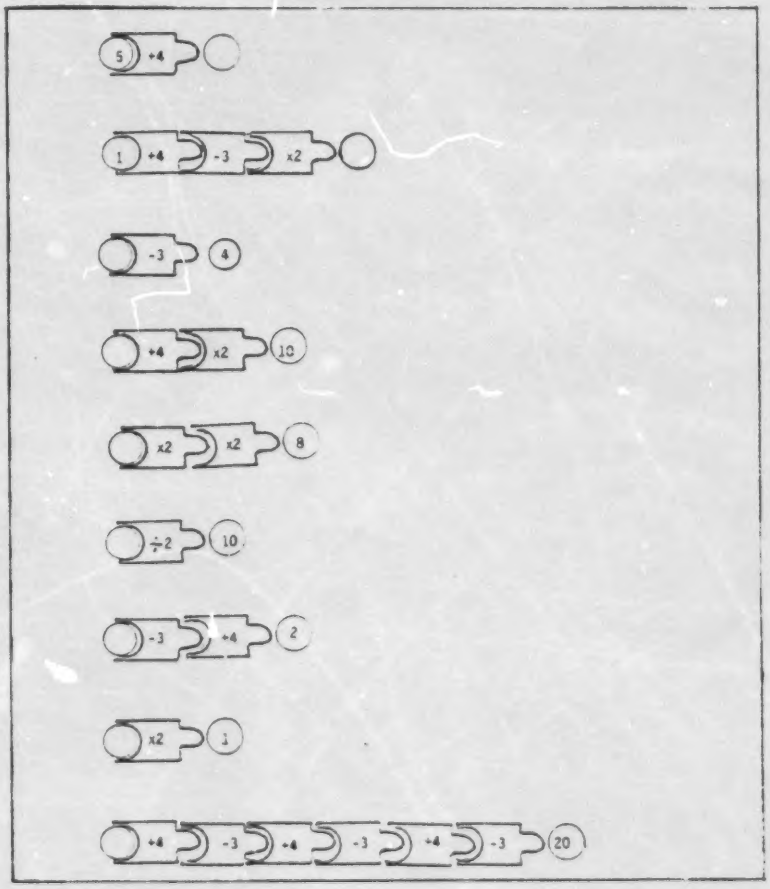
P3 is different. This time some number went into a +4 machine and the result was 6. What number went into the machine to give 6? Right 2; 2 + 4 = 6, so 2 was the number. You put 2 in the circle.

Sometimes you have to figure out what the result is; for other questions you have to figure out what number we started with - like the one we just did.



WORKING TIME:

Untimed, but usually required about 5 minutes.



SCALE #20: Boxes

NOTE: This scale contained 3 pages, each of which was preceded by rather lengthy instructions and each of which was separately timed.

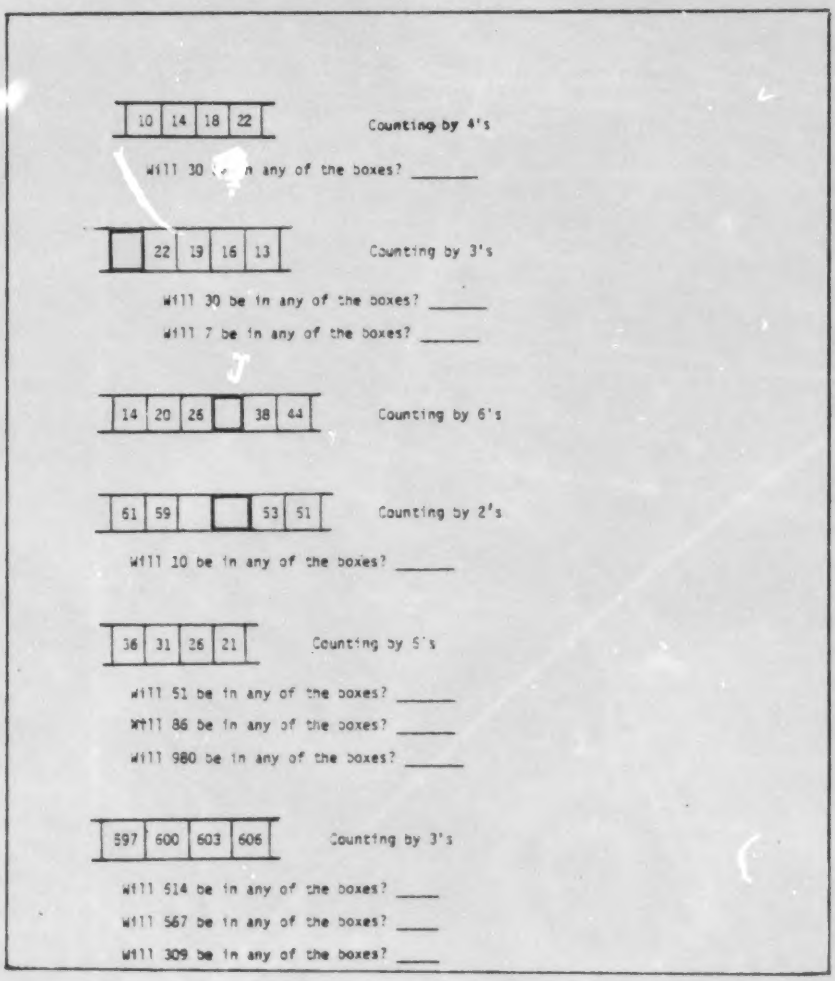
WORKING TIME: Untimed, but usually required about 4 minutes for each of the 3 pages.

ABBREVIATED DIRECTIONS (First page):

Look at the first row of boxes. It says, "Counting by 4's" and you can see there is always a difference of 4 between the numbers. Below the boxes it says, "Will 30 be in any of the boxes?" If we are counting by 4's, what would go in the next box after 22? Right, 26; 22 + 4 = 26. What comes next? Right, 30; So 30 will be in one of the boxes, so you write YES in the blank.

Look at the next row of boxes. One of the boxes has a heavy line around it. That means we have to figure out what goes in that box. If we are counting by 3's what comes next? Right, 22 + 3 is 25, so you put 25 in the box.

Then it says "Will 30 be in any of the boxes?" Let's keep going from 25; we have 13, 16, 19, 22, 25; what comes next? Right 28. And what would come next? Right 31. So 30 would not be in any of the boxes, so you put NO in the blank.


10 14 18 22
Counting by 4's
Will 30 be in any of the boxes? _____
22 19 16 13
Counting by 3's
Will 30 be in any of the boxes? _____
Will 7 be in any of the boxes? _____
14 20 26 38 44
Counting by 6's
61 59 53 51
Counting by 2's
Will 10 be in any of the boxes? _____
36 31 26 21
Counting by 5's
Will 51 be in any of the boxes? _____
Will 86 be in any of the boxes? _____
Will 980 be in any of the boxes? _____
597 600 603 606
Counting by 3's
Will 514 be in any of the boxes? ____
Will 567 be in any of the boxes? ____
Will 309 be in any of the boxes? ____

6.1

SCALE #20: Boxes (con't)

ABBREVIATED DIRECTIONS (Second page)

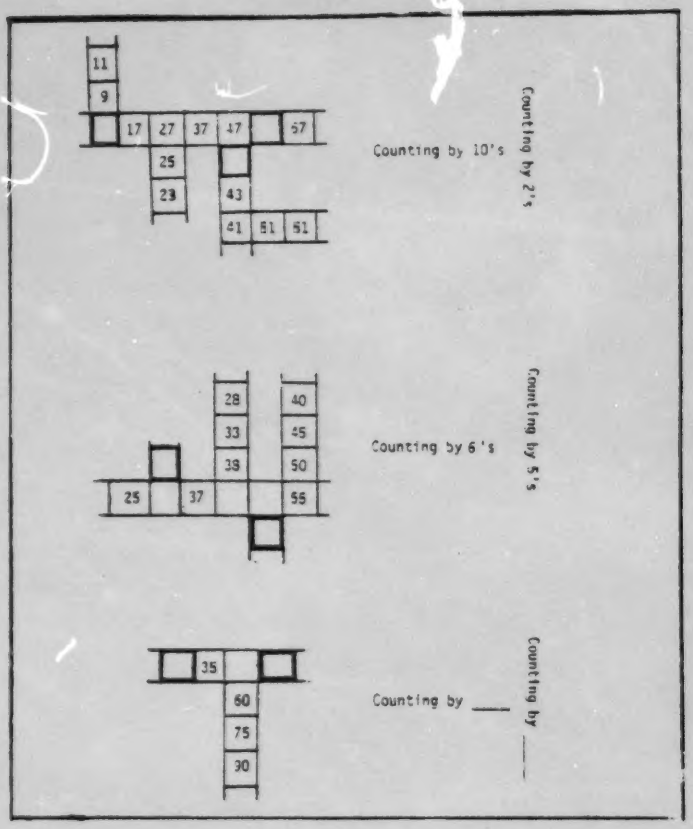
On this page we can be counting by one number going across and by a different number going up and down.

When we go across it says we are "counting by 10's".

It also says "counting by 2's". It's written sideways and that means whenever we are going up and down in this picture we are counting by "2's".

You have to fill in the heavy boxes on this page.

The bottom picture is harder because you have to do 2 things. First, figure out for yourself how much we are counting by and write the answers in the blanks; then fill in the heavy boxes.

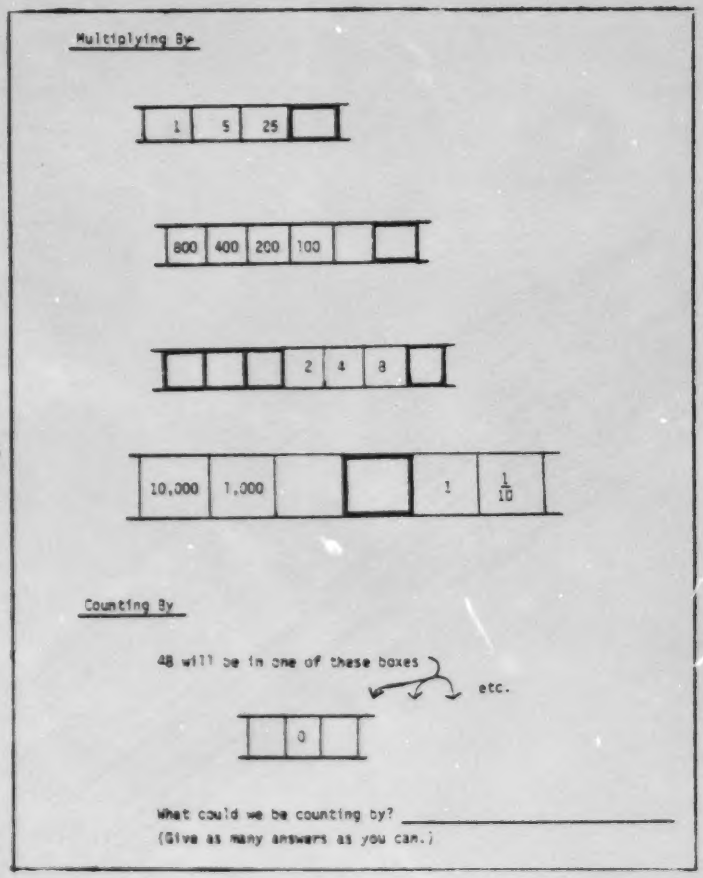


ABBREVIATED DIRECTIONS (Third page)

At the top of the page it says: "Multiplying By" Instead of always counting by a certain number we will be multiplying each time.

Look at the first one. We have 1, we multiply the 1 by a certain number to get 5, then we multiply the 5 by a certain number to get 25. What are we multiplying by each time? Right, 5; so 5 x 25 = what? Right, 125. So you put 125 in the box.

At the bottom of the page it says: "Counting By". When you get to that one you won't be multiplying any more, you'll be counting again - like the last page. The answer blank for that one is quite long. That's because you're supposed to give as many correct answers as you can for that problem.



SCALE #21: Place Value

ABBREVIATED DIRECTIONS:

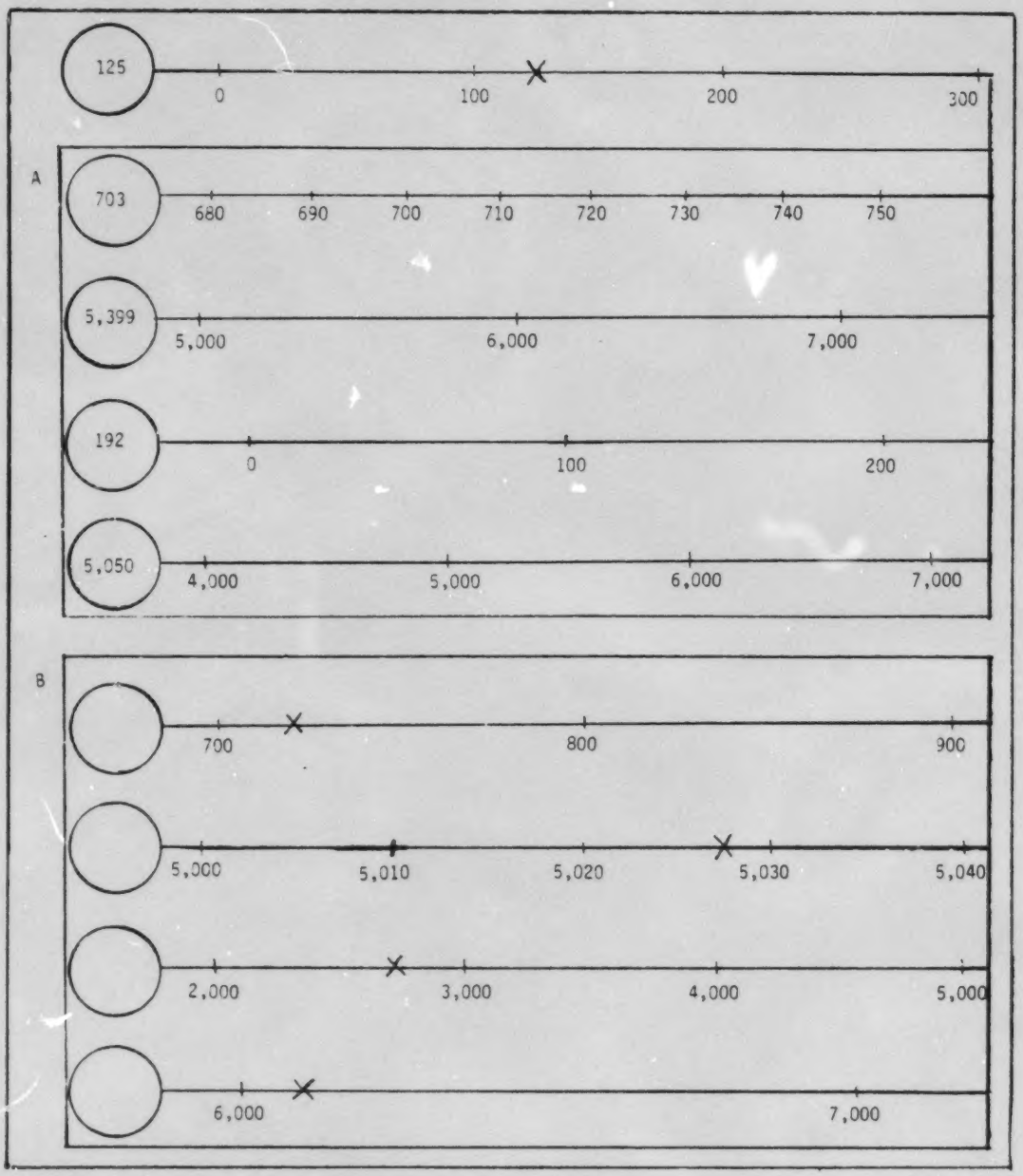
Look at the number line at the top of the page.

At the left side there is a circle with "125" written inside. Then there is an "x" on the number line to show where 125 is. The "x" is not just anywhere between the 100 and 200 mark - it's about where 125 actually would be.

In Box A, you figure out where the number would go on the number line and then mark an x on that spot.

In Box B, you figure out what number goes with the "x", and write that number in the circle.

WORKING TIME: Untimed, but usually required about 4 minutes.

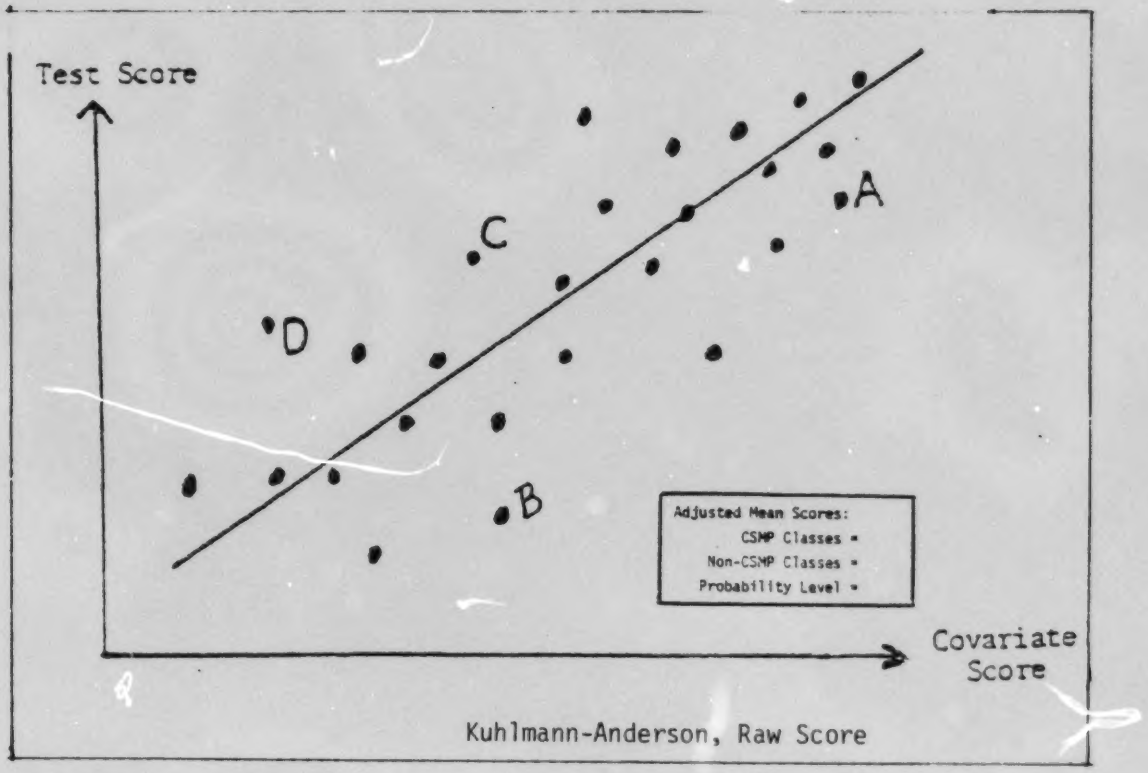


Appendix B

Scale-by-Scale Statistical Analysis

This Appendix contains information on each MANS scale. There are two pages for each scale. The left hand page provides numerical data, especially item analyses, based on a summary across students. The right hand page is a graph showing mean scores of each class.

Right Hand Pages



The right hand pages are for the graphs of class means. Each CSMP class is represented by an x and each Non-CSMP class by a dot (•). Based on this set of class means thus graphed, the regression line has been drawn. This line is the best linear prediction of mean class test score that can be made from knowing the ability level of the class. Note in the example that classes A and B fall well below the regression line, or are scoring well below what would be predicted for them form knowing the ability level of the class, while classes C and D fall well above the regression line. Note also that, although class A had a slightly higher mean score on the test than did class C, class C did much better given the relative ability scores of the two classes When the class means generally fall close to the regression line, test scores are well predicted by the covariate; when they are more dispersed from this line, the covariate is a less effective predictor.

The mean scores across CSMP and across Non-CSMP classes, adjusted for Kuhlmann-Anderson scores, are also given together with the p-value obtained from the F-test. (The p-value is the probability that a difference in mean scores that large could have occurred by chance alone. Hence, the smaller the p-value (especially below, say, .05), the more likely it is that there are "real" differences between CSMP and Non-CSMP classes.)

. Left Hand Pages

Test Items	Percent Correct		Biserial	Common Wrong Answers
	CSMP	Non-CSMP		
①	②		③	④
Number of Students	⑤			
KR 20 Reliability Homogeneity	⑥			
Correlation with Kuhlmann-Anderson	⑦			
Mean raw score " "	⑧			

⑨ Mean Ability Level:

	1	2	3	4	All	sd
CSMP						
Non-CSMP						

⑩ Notes

① The test items listed here are shown in a greatly compressed and sometimes altered form so that what required a full page on the student copy of the test (See Appendix A) can be squeezed into this space.

② These are the percentages of students getting the item correct.

③ This is the r-biserial correlation coefficient between the item and the score on the scale (excluding the item from the score). A large biserial (above say, .8) means a high correspondence between how well students did on this item and how well they did on the scale as a whole. A low biserial (below, say, .3) means a lack of correspondence.

④ On some scales, wrong answers given by more than 5% of the students are listed here. When more than 20% of the students give this response that percentage is given in parenthesis after the wrong response.

⑤ The number of CSMP students taking any scale varies from about 80 to 170 (the lower figure occurring if the scale was given to only a random half of each class). For Non-CSMP, these numbers were a little lower.

⑥ The KR 20 reliability coefficient in a measure of the degree to which the items in a scale are testing a single underlying ability. A large KR 20 (above say, .85) means a correlation among the items; a low KR 20 (below say .6) means a low correlation among the items and not a single underlying ability.

⑦ The Kuhlmann-Anderson ability tests were administered to all students earlier in the year. The score use used here is actually the total KA score minus the scores on the two math subtests of the KA.

(8) The mean raw score on the KA favored CSMP students by fro[illegible] out 1 to 3 raw score units (the standard deviation for the KA being around 17). The effect of this advantage on the item percents correct ((2)) depends on several things but typically it might amount to about 2-4%. Thus a difference of 4% correct in favor of CSMP students might be entirely accounted for by the CSMP students being slightly higher in general academic ability.

(9) Students were allocated to quartiles based on their KA scores in such a way that roughly equal numbers of students were in each quartile. Because of above average ability of the total group of students the "break" points between the lowest and the second lowest quartiles was not at the 25th percentile but at the 50th. The other two break points were at about the 70th and 80th percentile. The mean KA score within quartiles was almost the same for CSMP and Non-CSMP except than in the lowest quartile, CSMP students had a mean score of about 5 points higher than Non-CSMP.

(10) The notes point out or amplify some of the more important or interesting data. One of the things given in the notes is the percent correct for those items which were also administered last year to basically the same group of students who were then in third grade.

#1 - Computation-Addition

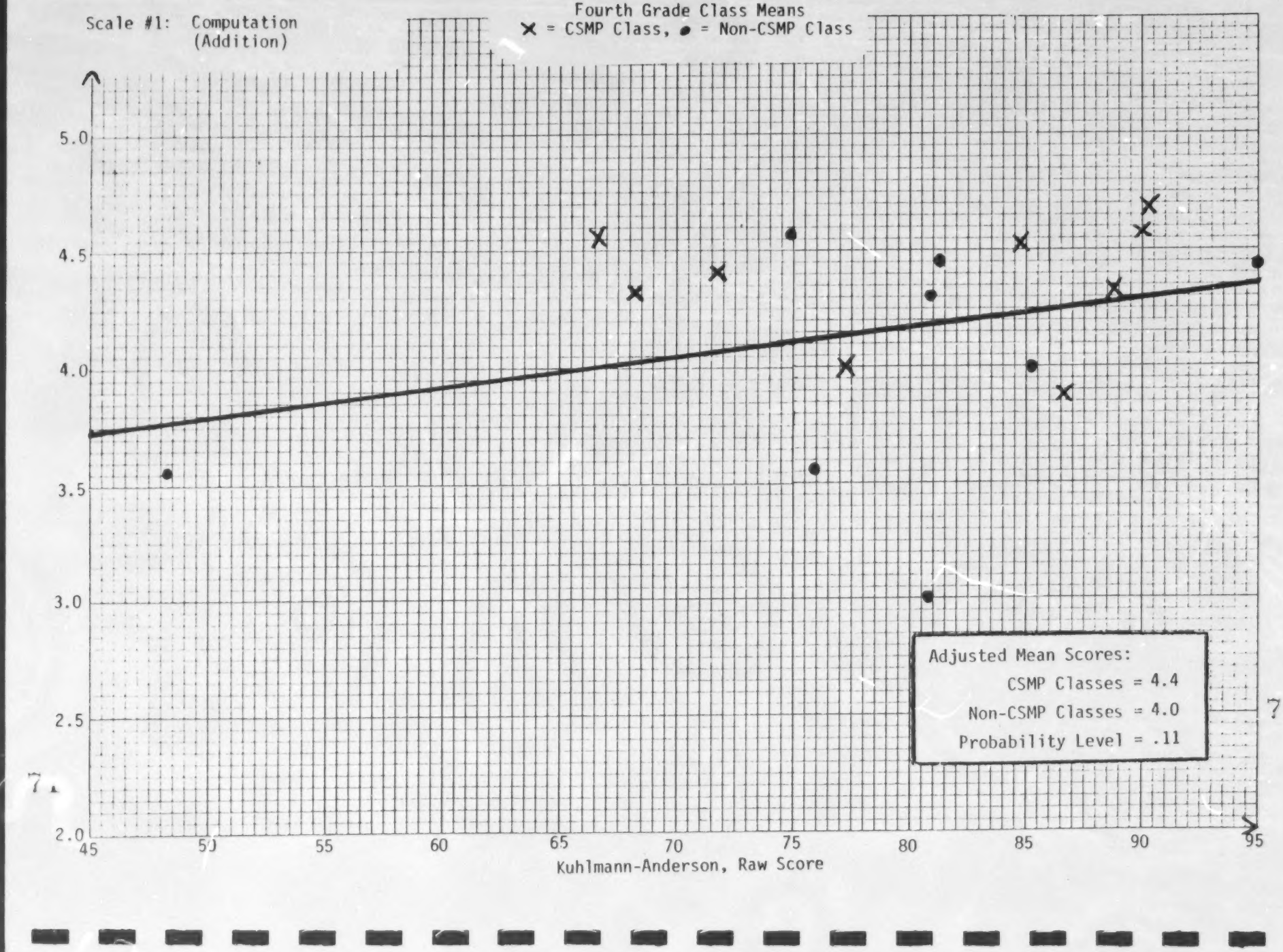
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
347 + 128	89	94	.12
708 + 293	90	91	.35
793 + 8 = □	89	78	.22
574 + 258 = □	88	81	.56
$43.57 6.89 .62 7.15	74	54	.47
Number of Students	84	68	
KR 20 Reliability Homogeneity	.24	.56	
Correlation with Kuhlmann-Anderson	.00	.29	
Mean raw score, " "	79.8	76.3	

Means by Ability Level:

	1	2	3	4	All	sd
CSMP	4.3	4.5	4.1	4.6	4.3	0.8
Non-CSMP	3.6	3.7	4.8	4.2	4.0	1.1

Note:

1. As expected, this test was fairly easy for most students (about 3/4 of the students made no more than 1 error) and for most classes (14 of the 18 classes averaged at least 4 out of 5 correct). Thus most classes have mastered these addition skills and the Kuhlmann-Anderson is not a particularly effective predictor at either the student or class level; in fact for CSMP students the correlation between this scale and the Kuhlmann-Anderson is 0.


Scale #1: Computation (Addition)
Fourth Grade Class Means
× = CSMP Class, ● = Non-CSMP Class
5.0
4.5
4.0
3.5
3.0
2.5
2.0
45
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 4.4
Non-CSMP Classes = 4.0
Probability Level = .11

#2 - Computation-Subtraction

Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
606 -482	79	76	.83
758 -169	82	78	.58
203 - 8 = □	74	65	.66
680 - 155 = □	68	69	.65
1024 -727	65	60	.74
Number of Students	82	68	
KR 20 Reliability Homogeneity	.78	.71	
Correlation with Kuhlmann-Anderson	.50	.54	
Mean raw score, " "	78.7	77.8	

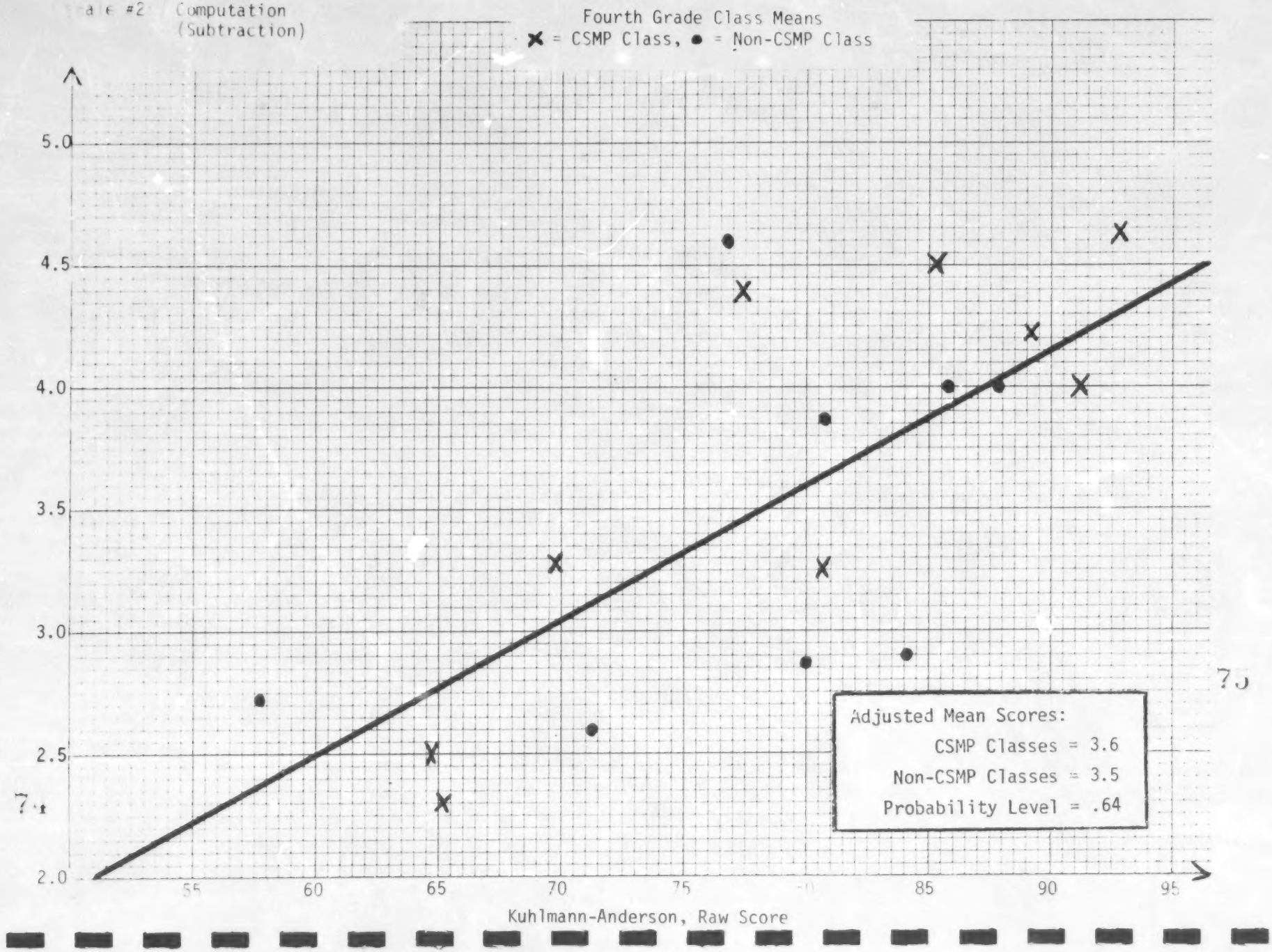
Means by Ability Level:

	1	2	3	4	All	sd
CSMP	2.7	3.5	4.3	4.2	3.7	1.6
Non-CSMP	1.9	3.8	3.7	4.1	3.5	1.5

Note:

1. While most students (more than 60%) got at least 4 of the 5 items correct, about 20% of the students did not get more than 2 items correct. Thus fourth graders in general are not anywhere near the level of mastery in subtraction that they are in addition. In some ways they are better at the multiplication algorithm (see Scale #3) than the subtraction algorithm even though multiplication is taught later and is supposedly harder than subtraction.

Scale #2 Computation (Subtraction)



#3 - Computation-Multiplication

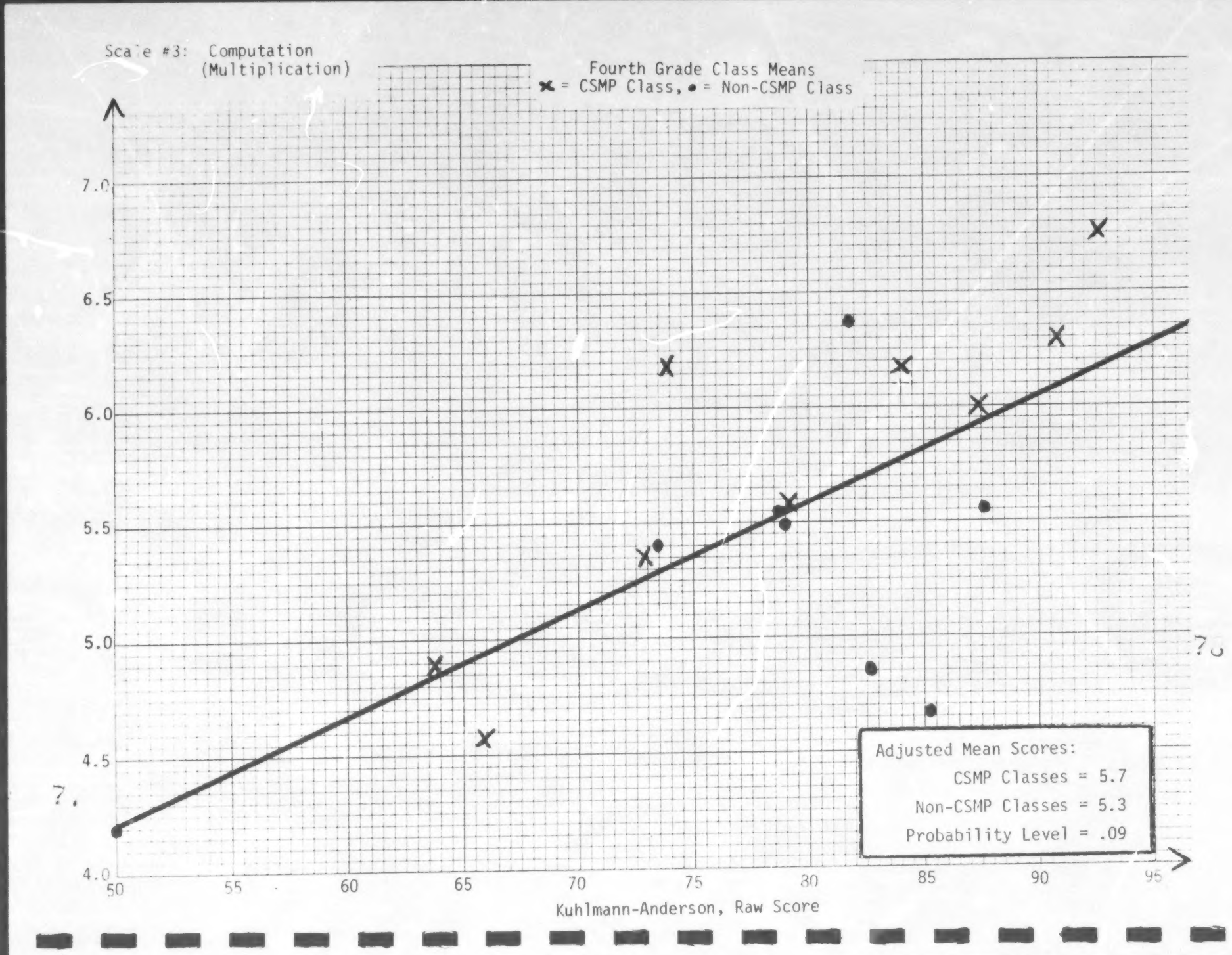
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
5 x 3 = ☐	98	97	.87
3 x 35 = ☐	84	81	.62
4 x 210 = ☐	90	78	.71
x7	83	82	.77
43 x7	76	75	.72
623 x4	81	81	.76
314 x34	48	35	.41
Number of Students	80	68	
KR 20 Reliability Homogeneity	.76	.67	
Correlation with Kuhlmann-Anderson	.45	.59	
Mean raw score, " "	80.2	76.9	

Means by Ability level:

	1	2	3	4	All	sd
CSMP	5.0	5.1	6.2	6.6	5.6	1.7
Non-CSMP	4.2	5.4	6.0	5.4	5.3	1.6

Note:

1. Most of the difference in class mean scores can be accounted for by 2 low scoring Non-CSMP classes. This may also explain the relatively low mean score for high ability (Q_4) Non-CSMP students.


Scale #3: Computation
(Multiplication)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
7.0
6.5
6.0
5.5
5.0
4.5
4.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 5.7
Non-CSMP Classes = 5.3
Probability Level = .09

#4 - Computation-Division

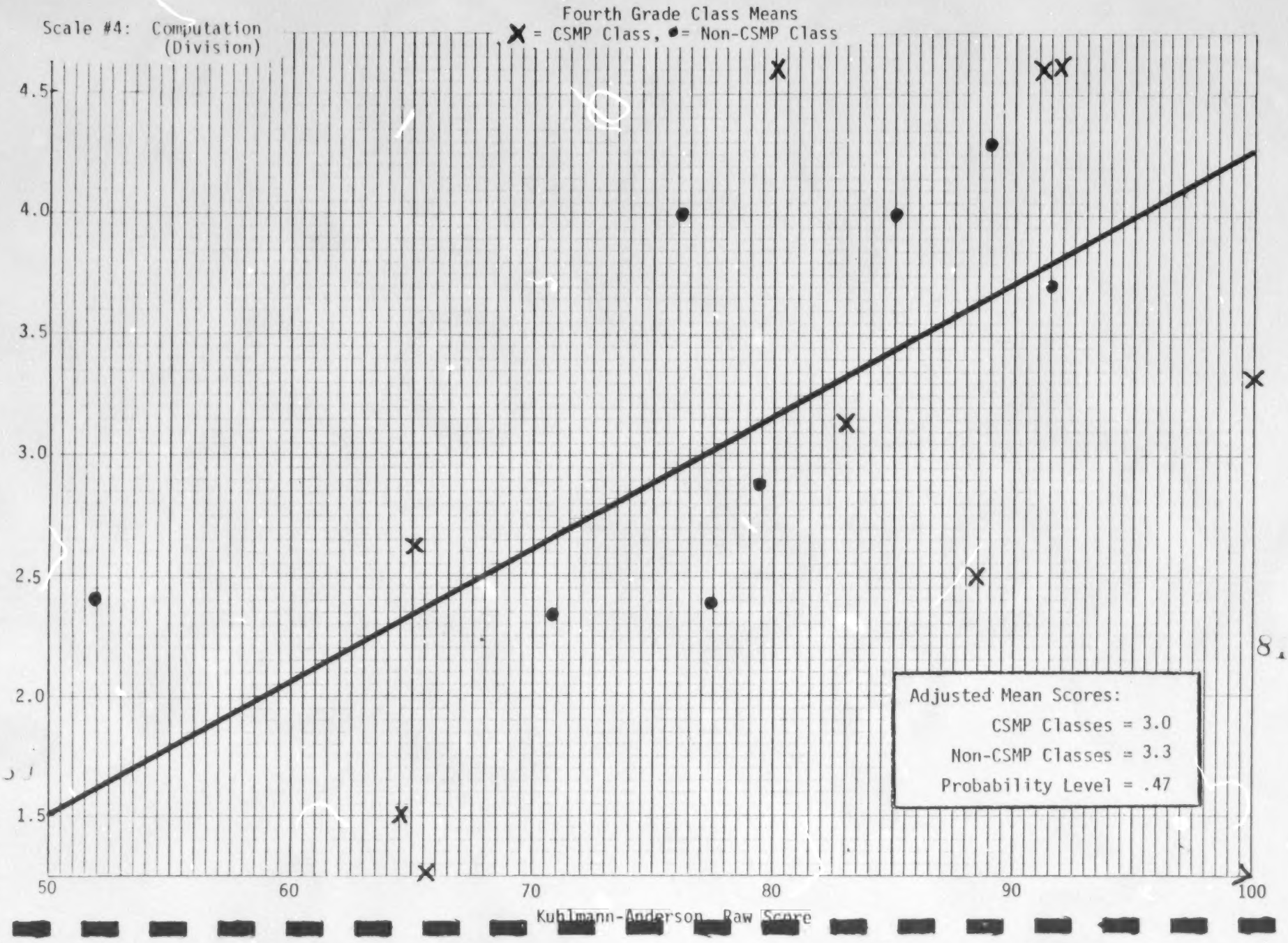
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
$30 \div 6 = \square$	80	90	.55
$219 \div 3 = \square$	46	42	.70
$144 \div 6 = \square$	43	27	.74
$7\overline{)56}$	67	73	.51
$4\overline{)128}$	56	60	.80
$6\overline{)94}$	46	55	.73
Number of Students	81	73	
KR 20 Reliability Homogeneity	.80	.72	
Correlation with Kuhlmann-Anderson	.55	.60	
Mean raw score, " "	80.0	77.4	

Means by Ability Level:

	1	2	3	4	All	sd
CSMP	1.4	3.9	3.4	4.3	3.4	2.0
Non-CSMP	2.0	2.9	3.6	5.0	3.5	1.7

Note:

1. The first 3 items used the symbol " $\div$ ", which was more frequently used in CSMP classes; the last 3 items used "$\overline{)}$", which was more frequently used in Non-CSMP classes. These differences are reflected somewhat in the item statistics. Compare the third and the sixth items.
2. It is apparent that there is a wide variation in how well fourth graders do division problems. About 20% of the students did not attempt more than 2 items, yet 20% also got all 6 items correct.
3. This the only scale on which Non-CSMP classes did better than CSMP classes, though the difference was small and not significant.


Fourth Grade Class Means
X = CSMP Class, • = Non-CSMP Class
Scale #4: Computation (Division)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
50
60
70
80
90
100
Kuhlmann-Anderson Raw Score
Adjusted Mean Scores:
CSMP Classes = 3.0
Non-CSMP Classes = 3.3
Probability Level = .47

SCALE #5 - Word Problems-One Step

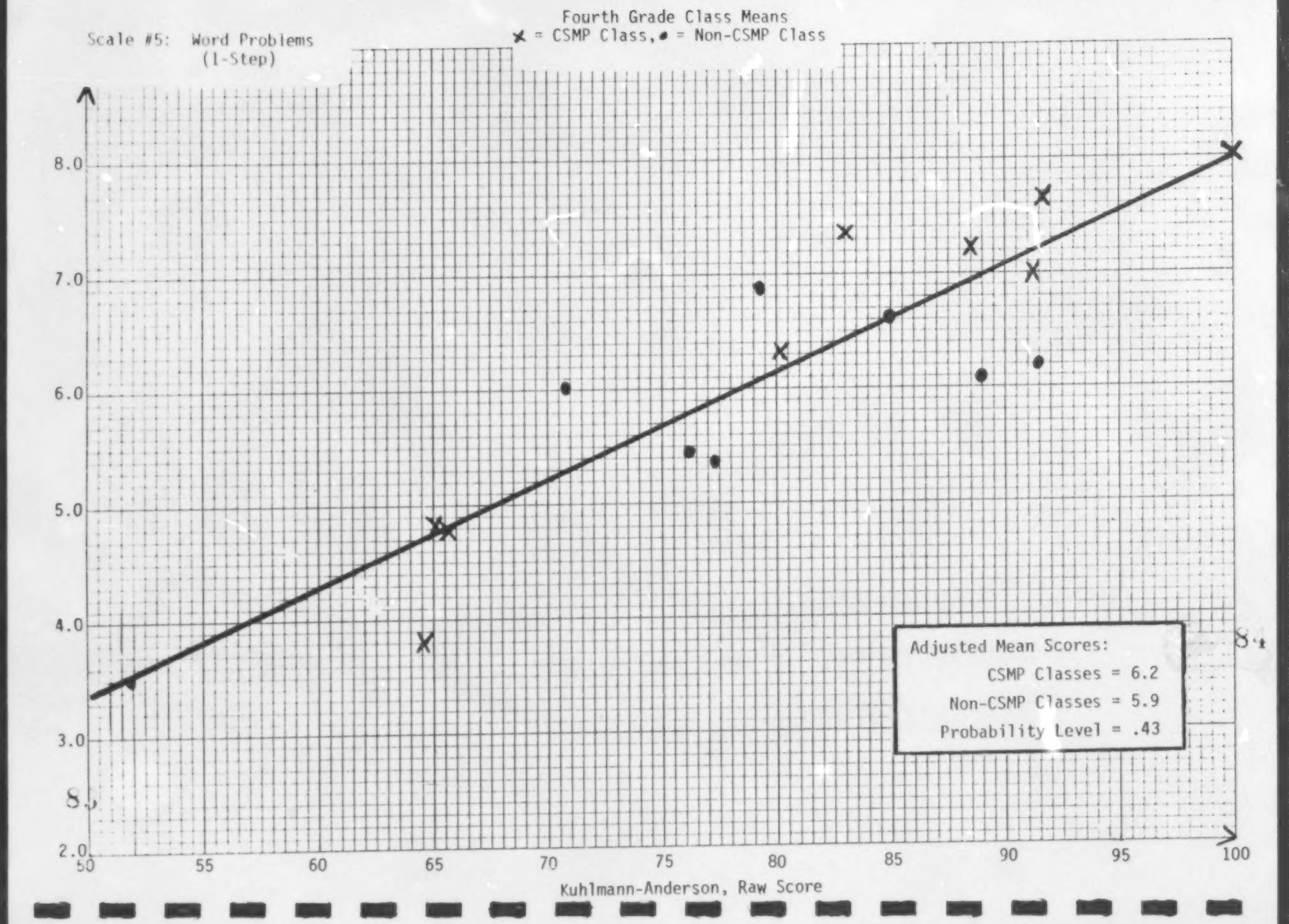
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
1. Bill started with 22 candy bars. He sold 17 of them. How many candy bars did he have then? _____	91	84	.75	
2. Andy bought 4 cartons of cola. Each carton holds 8 bottles. How many bottles did he buy altogether? _____	81	78	.89	12
3. Jane has 25¢. Pencils cost 7¢ each. How many pencils can Jane buy?_____ How much money will she have left over? _____	75 71	73 73	.84 .99	18
4. Ann has 31¢. She wants to buy a game that cost 45¢. How much more money does she need?_____	86	77	.70	
5. Joe started with 40 marbles. His father gave him 8 more. Then his mother gave him 9 more. How many marbles did he have then?_____	91	82	.88	
6. There are 4 men who take turns parking cars. Altogether they will park 24 cars. How many will each man park?_____	65	71	.67	28
7. Tom has 3¢ more than Ann Tom has 5¢ less than John. If John has 20¢, how much does Ann have?_____	46	36	.62	
Number of Students	81	73		
KR 20 Reliability Homogeneity	.82	.83		
Correlation with Kuhlmann-Anderson	.64	.51		
Mean raw score, " "	80.0	77.4		

Means by Ability Level:

	1	2	3	4	All	sd
CSMP	3.9	6.4	6.5	7.7	6.1	2.1
Non-CSMP	3.7	6.1	6.3	6.9	5.7	2.3

Note:

1. These straight forward items were fairly well done (except for the last quesiton which was more difficult); over half the students got at least 7 out of 8 questions correct.


Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
Scale #5: Word Problems (1-Step)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
50
55
60
65
70
75
80
85
90
95
100
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 6.2
Non-CSMP Classes = 5.9
Probability Level = .43

SCALE #6 - Word Problems-Two Step

Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
1. On Saturday Amy and Susan made $13 selling lemonade. On Sunday they made $5. They put their money together and divided it evenly. How much did each girl get? _____	63	50	.51	18
2. Peter bought 4 blue pencils and 3 red pencils. Each pencil costs 10¢. What was the total cost? _____	83	82	.60	
3. Pam's allowance is $1.00 each week. She always spends 20¢ and saves the rest. How much will she save in 5 weeks? _____	55	40	.67	
4. There are 40 apples in my barrel now. I will eat 2 apples every day. How many apples will be left in my barrel after 5 days? _____	69	54	.73	
5. Jim has $10 in his bank now. Each week he will add $5 to his bank. In how many weeks will he have $30 in his bank? _____	55	38	.48	6.3
6. Joel has 20 sunflower seeds for his pet birds. Each bird gets the same number of seeds; no seeds are left over. How many pet birds could Joel have? _____ (List all the possible answers.)	63	32	.54	
Number of Students	84	68		
KR 20 Reliability Homogeneity	.74	.69		
Correlation with Kuhlmann-Anderson	.57	.66		
Mean raw score, " "	79.8	76.3		

Means by Ability Level:

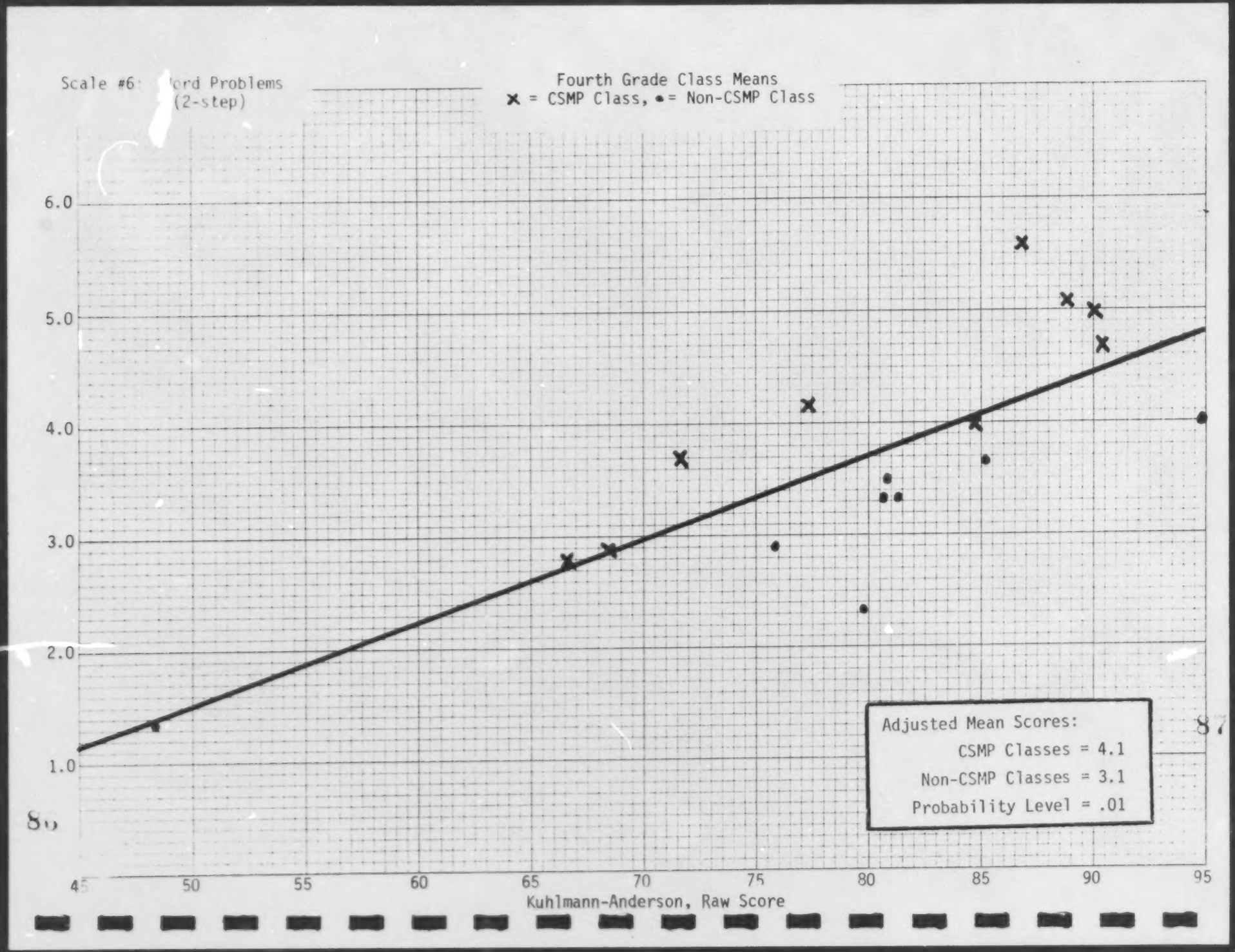
	1	2	3	4	All	sd
CSMP	2.5	3.3	4.3	5.2	3.9	1.5
Non-CSMP	1.5	3.0	3.3	3.9	3.0	1.8

Note:

1. This was the only word problem scale on which the CSMP class means were significantly higher than the Non-CSMP class means.

2. Items 2, 4 and 5 were administered last year when these students were third graders. The mean percent correct on these 3 items was 62 for CSMP students and 48 for Non-CSMP this year versus 54 and 44 last year.

3. On the last question there were 6 possible correct answers (1,2,4,5,10 and 20). A response was counted correct if 4 or more of these were given. The least frequently given responses were "1" (given by 24% of the students) and "20" (given by 45%). The other 4 responses were each given by about 74% of the CSMP students and 49% of the Non-CSMP students.


Scale #6: ord Problems
(2-step)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
6.0
5.0
4.0
3.0
2.0
1.0
45
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 4.1
Non-CSMP Classes = 3.1
Probability Level = .01

#7 - Word Problems-Hits and Misses

Test Items*	Percent Correct		Biserial Correlation	Common Errors
	CSMP	Non-CSMP		
If we hit the target, we go up 5 points: If we miss the target we go down 2 points: Bill started with a score of 0. Then he had 2 hits and 3 misses. What was his score then? _____	43	21	.87	
Mary started with a score of 7 below zero. Her next 2 throws were hits. What was her score then? _____	37	19	.83	
John started with a score of 0. Then he had 1 hit and 4 misses. What was his score then? _____	37	24	.77	1(20%)
Toni started with a score of 9 below zero. Her next 4 throws were misses. What was her score then? _____	26	22	.95	
Hank started with a score of 0. Then he had 3 throws. What could his score be then? _____ (List all the possible answers.)	27	12	.79	
Number of Students	82	82		
KR 20 Reliability Homogeneity	.81	.83		
Correlation with Kuhlmann-Anderson	.42	.49		
Mean raw score, " "	78.7	77.8		

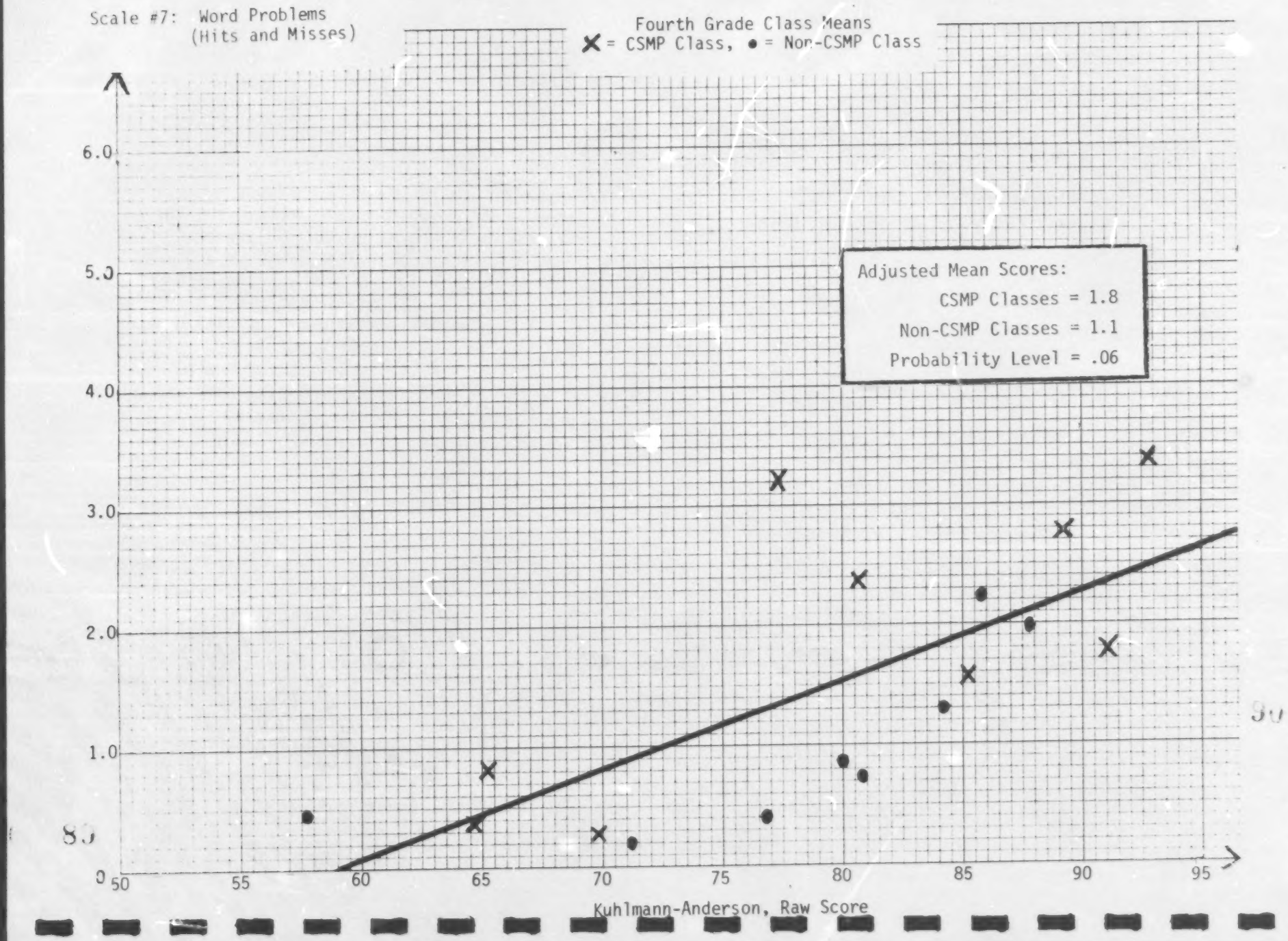
Means by Ability Level:

	1	2	3	4	All	sd
CSMP	0.6	1.4	2.1	2.6	1.7	1.8
Non-CSMP	0.2	0.5	0.9	2.4	1.0	1.5

Note:

1. Many students (40% of CSMP, 62% of Non-CSMP) apparently did not understand the problem and got none of the 5 items correct (although they attempted to answer most of the questions).
2. The possible answers to the last question were 15, 8, 1 and 6 below zero. Three or more of these answers were necessary for a response to be counted as correct.

*Hits and Misses" is explained more fully on student copy, Page A8.


Scale #7: Word Problems
(Hits and Misses)
Fourth Grade Class Means
✕ = CSMP Class, ● = Non-CSMP Class
Adjusted Mean Scores:
CSMP Classes = 1.8
Non-CSMP Classes = 1.1
Probability Level = .06
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score

SCALE #8: Word Problems - Miscellaneous

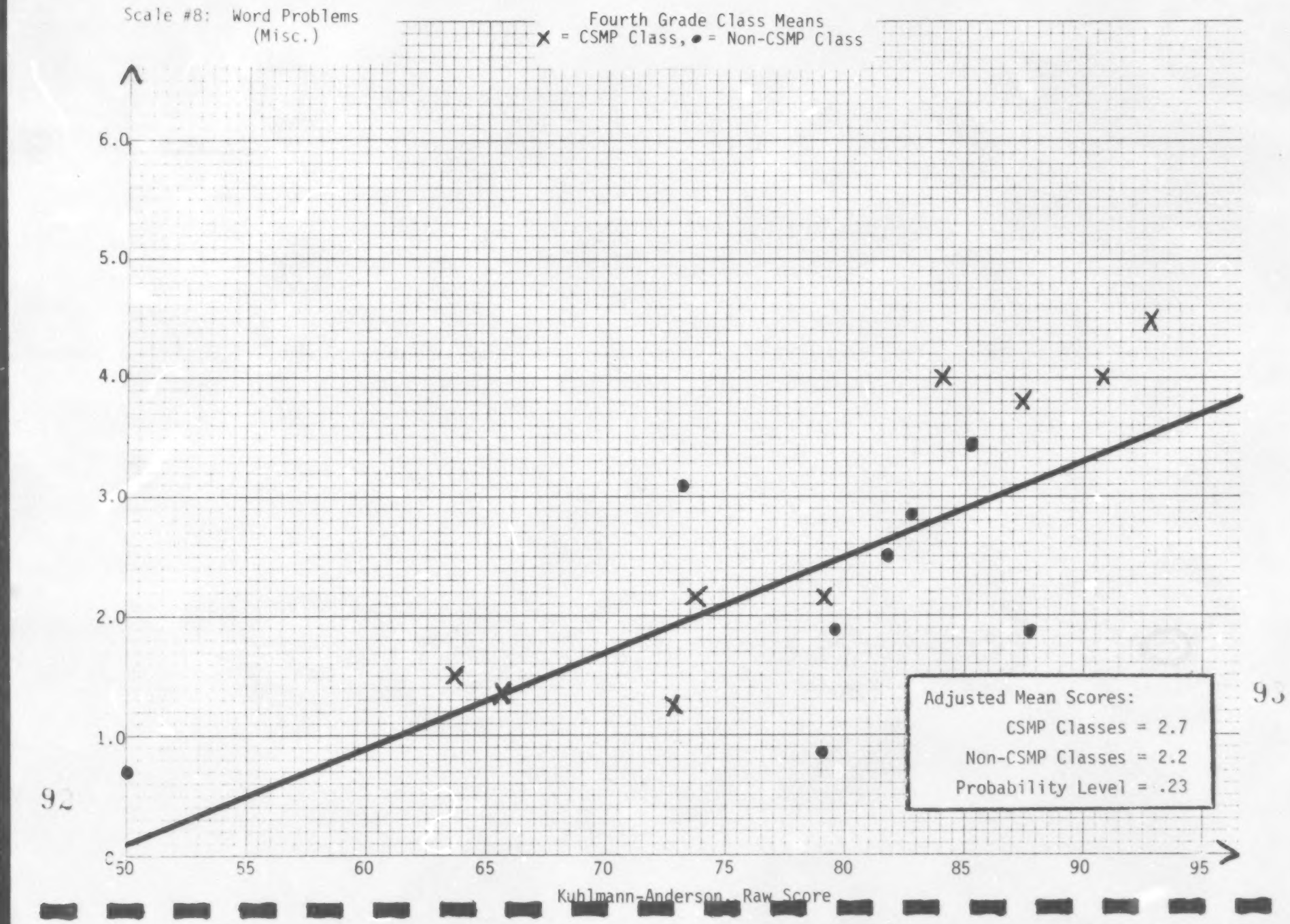
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
Mr. Jones owns a truck company. His trucks can carry 3 cars each. He is loading 14 cars into trucks. How many trucks will he need? _____	53	47	.53	4
It takes 4 men to keep an oil well working. There are 18 men ready to work. How many oil wells can they keep working? _____	55	49	.56	5
We buy 3 pounds of nuts for 25¢. How many pounds can we buy for $1.00? _____	51	53	.66	4
There are 70 ounces in 5 cans of cola. How many ounces are there in 3 cans? _____	28	18	.73	
Amy sold 10 tickets to her school dance. Altogether she got $3.50 for these tickets. How much did each ticket cost? ____	45	25	.80	
The meter on Bill's car shows 247 $\frac{3}{10}$ miles. He drives another $\frac{7}{10}$ of a mile. What will the meter show then? ____ (2 4 7 3 /)	40	25	.70	2470
Number of Students	80	68		
KR 20 Reliability Homogeneity	.82	.70		
Correlation with Kuhlmann-Andersor	.55	.47		
Mean raw score, " "	80.2	76.9		

Means By Ability Level:

	1	2	3	4	All	sd
CSMP	1.4	1.7	3.5	4.5	2.7	2.1
Non-CSMP	0.8	2.0	2.4	3.3	2.2	1.7

Note:

1. This scale consisted of 3 pairs of similar items. The first pair (items 1 and 2) dealt with dividing and rounding the answer (either up or down) and the second pair of questions (items 3 and 4) concerned ratios; on these two types there was only a slight advantage to CSMP. The third pair of questions involved decimals and CSMP students did much better.


Scale #8: Word Problems (Misc.)
Fourth Grade Class Means
X = CSMP Class, • = Non-CSMP Class
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson Raw Score
Adjusted Mean Scores:
CSMP Classes = 2.7
Non-CSMP Classes = 2.2
Probability Level = .23

SCALE #9 - Large Number Computation (+, ÷)

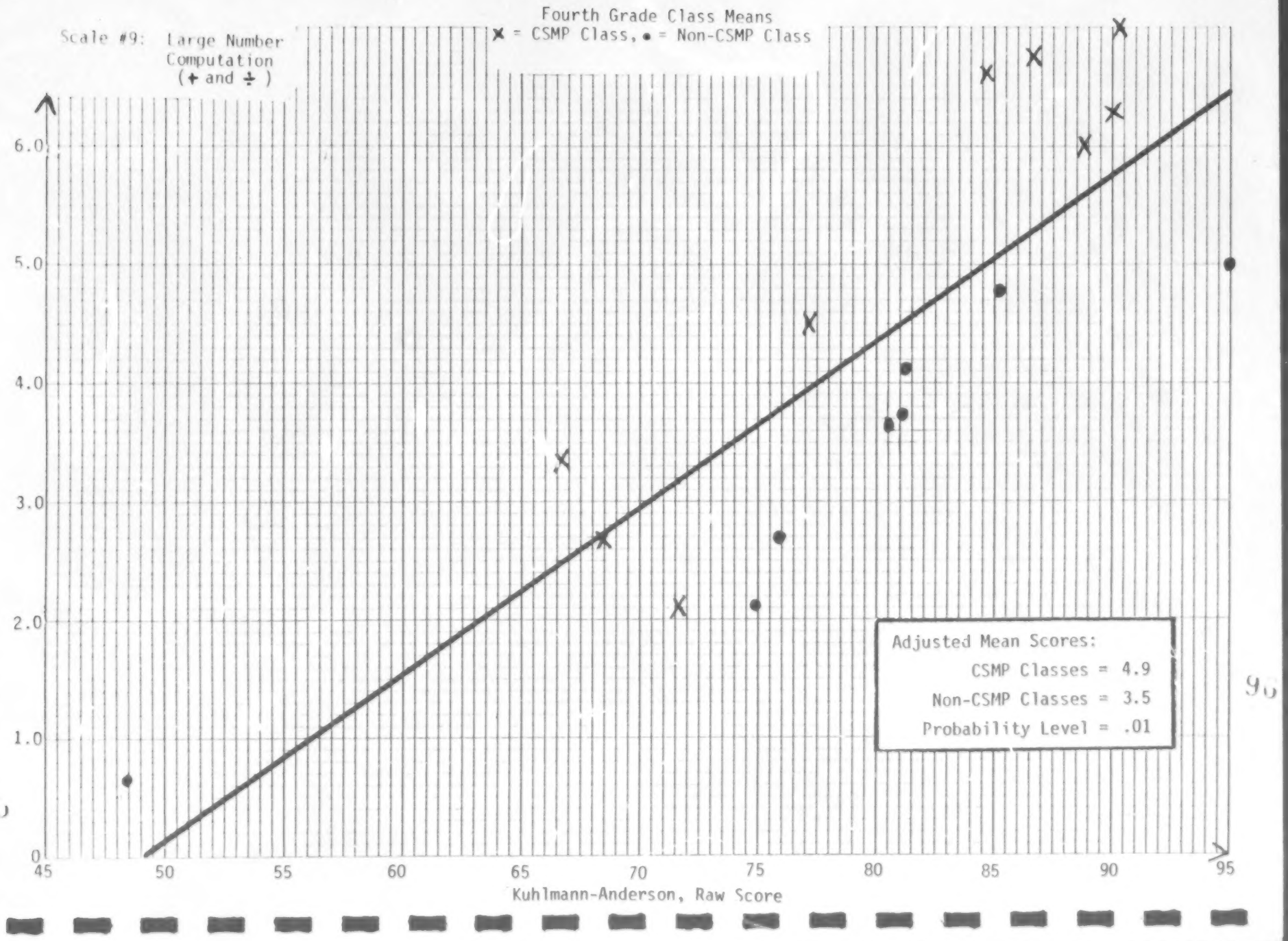
Test Items	Percent Correct		Biserial Correlation
	CSMP	Non-CSMP	
375 + ☐ = 600	61	44	.84
☐ + 125 = 250	76	65	.66
989 + ☐ = 1,000	65	35	.71
500 ÷ 2 = ☐	68	32	.94
700 ÷ 10 = ☐	71	51	.78
300 ÷ 50 = ☐	42	31	.74
800 ÷ ☐ = 200	49	31	.78
☐ ÷ 3 = 120	39	16	.74
600 ÷ ☐ = 4	25	16	.78
Number of Students	84	68	
KR 20 Reliability Homogeneity	.85	.86	
Correlation with Kuhlmann-Anderson	.64	.66	
Mean raw score, " "	79.8	76.3	

Means By Ability Level:

	1	2	3	4	all	sd
CSMP	2.7	3.4	5.9	7.9	5.0	2.9
Non-CSMP	0.8	2.4	3.9	6.7	3.2	2.8

Note:

1. CSMP students continue to do better on these types of questions than do Non-CSMP students. This is true regardless of ability level and operation (note that the first three problems might more accurately be labelled "subtraction" since that operation is probably used to solve the problem). See also Scale #10, a similar scale using the "−" and "x" symbols.
2. About 1/4 of the students did not attempt the last 2 problems. In fact 16% of the Non-CSMP students did not attempt the last 4 items.


Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
Scale #9: Large Number Computation (+ and ÷)
6.0
5.0
4.0
3.0
2.0
1.0
0
45
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 4.9
Non-CSMP Classes = 3.5
Probability Level = .01

SCALE #10 - Large Number Computation (−, ×)

Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial
525 - □ = 225	66	60	.52
□ - 250 = 150	32	18	.56
3 x 125 = □	83	69	.74
2 x 200 = □	93	83	.78
□ x 4 = 400	71	65	.63
□ x 250 = 500	73	43	.75
5 x □ = 2500	52	40	.63
150 x □ = 600	48	37	
Number of Students	82	68	
KR 20 Reliability Homogeneity	.72	.76	
Correlation with Kuhlmann-Anderson	.70	.55	
Mean raw score, " "	78.7	77.8	

Means Ability By Level:

	1	2	3	4	all	sd
CSMP	3.2	5.1	5.6	6.9	5.2	1.9
Non-CSMP	1.8	4.7	4.3	5.3	4.1	2.1

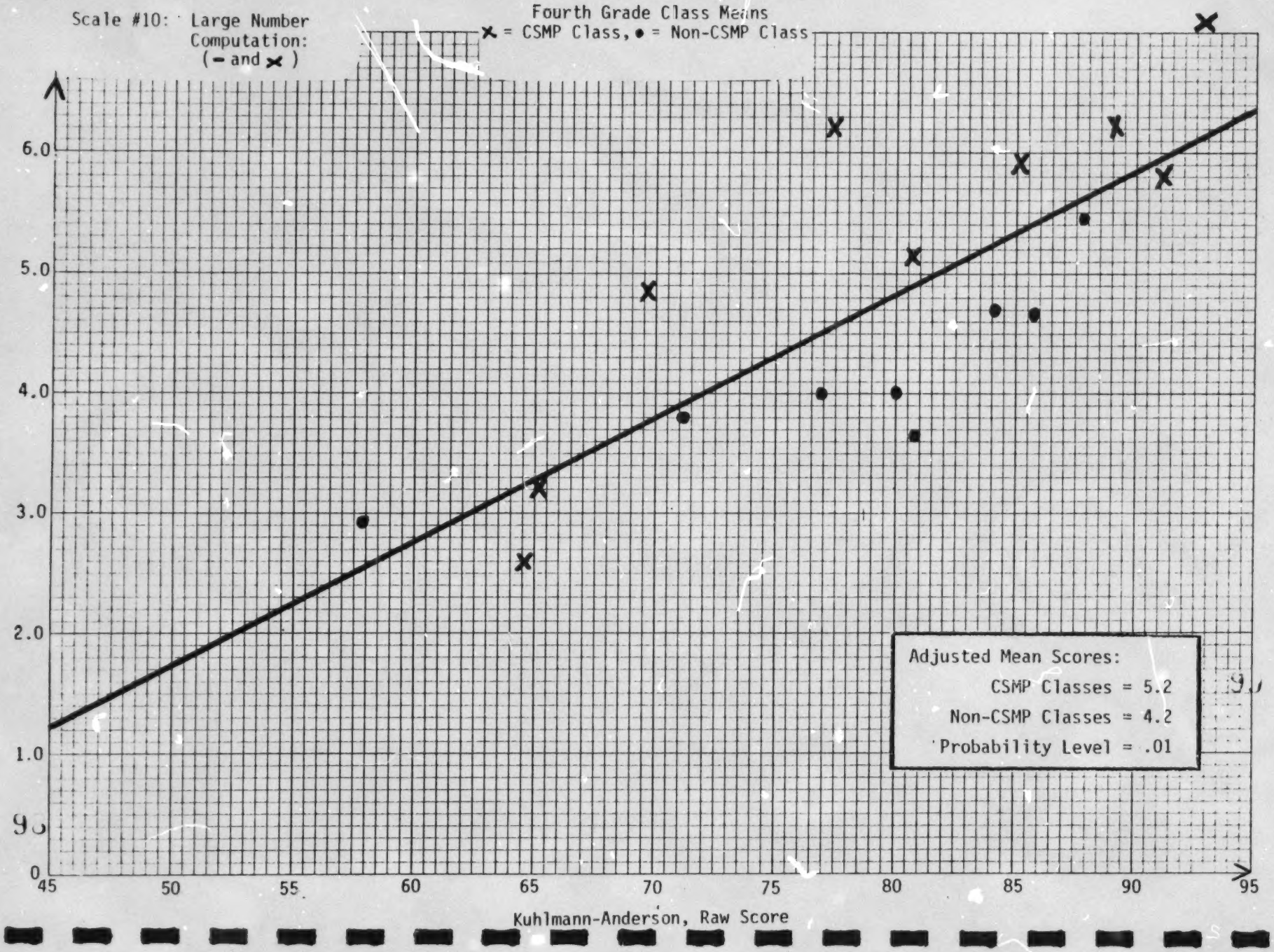
Note:

1. All of the items on this scale, plus the first 2 items of Scale #9, were also administered last year. The mean percent correct on those items this year was 66 for CSMP students and 52% for Non-CSMP students (versus 45 and 37 last year).

2. When one considers all the items in this scale and in Scale #9 the CSMP advantage was fairly consistent regardless of operation.

Mean Percent Correct, When Classified by:

	Given Sign of Operation (Number of Items)				Operation Needed to do Calculation (Number of Items)			
	+ (3)	− (2)	x (6)	÷ (6)	+ (1)	- (4)	x (3)	÷ (9)
CSMP	67	49	70	50	32	67	72	45
Non-CSMP	48	39	55	31	18	51	56	28


Scale #10: Large Number Computation: (− and ×)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
6.0
5.0
4.0
3.0
2.0
1.0
0
45
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 5.2
Non-CSMP Classes = 4.2
Probability Level = .01

SCALE #11 - Fractions

Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
$\frac{1}{3}$ of 15 = ☐	82	59	.69	
$\frac{1}{2}$ of 230 = ☐	66	36	.78	
$\frac{1}{2}$ of ☐ = 8	64	35	.76	4 (30%)
$\frac{1}{3}$ of ☐ = 12	49	15	.84	4 (30%)
$\frac{1}{2} + \frac{1}{2}$ = ☐	67	62	.73	$\frac{2}{4}$
$\frac{3}{5} - \frac{1}{5}$ = ☐	40	53	.61	$\frac{2}{10}$
$8\frac{1}{3} - 5$ = ☐	39	40	.65	
$1 - \frac{2}{3}$ = ☐	31	32	.56	
Number of Students	166	136		
KR 20 Reliability Homogeneity	.82	.81		
Correlation with Kuhlmann-Anderson	.65	.58		
Mean raw score, " "	79.3	77.1		

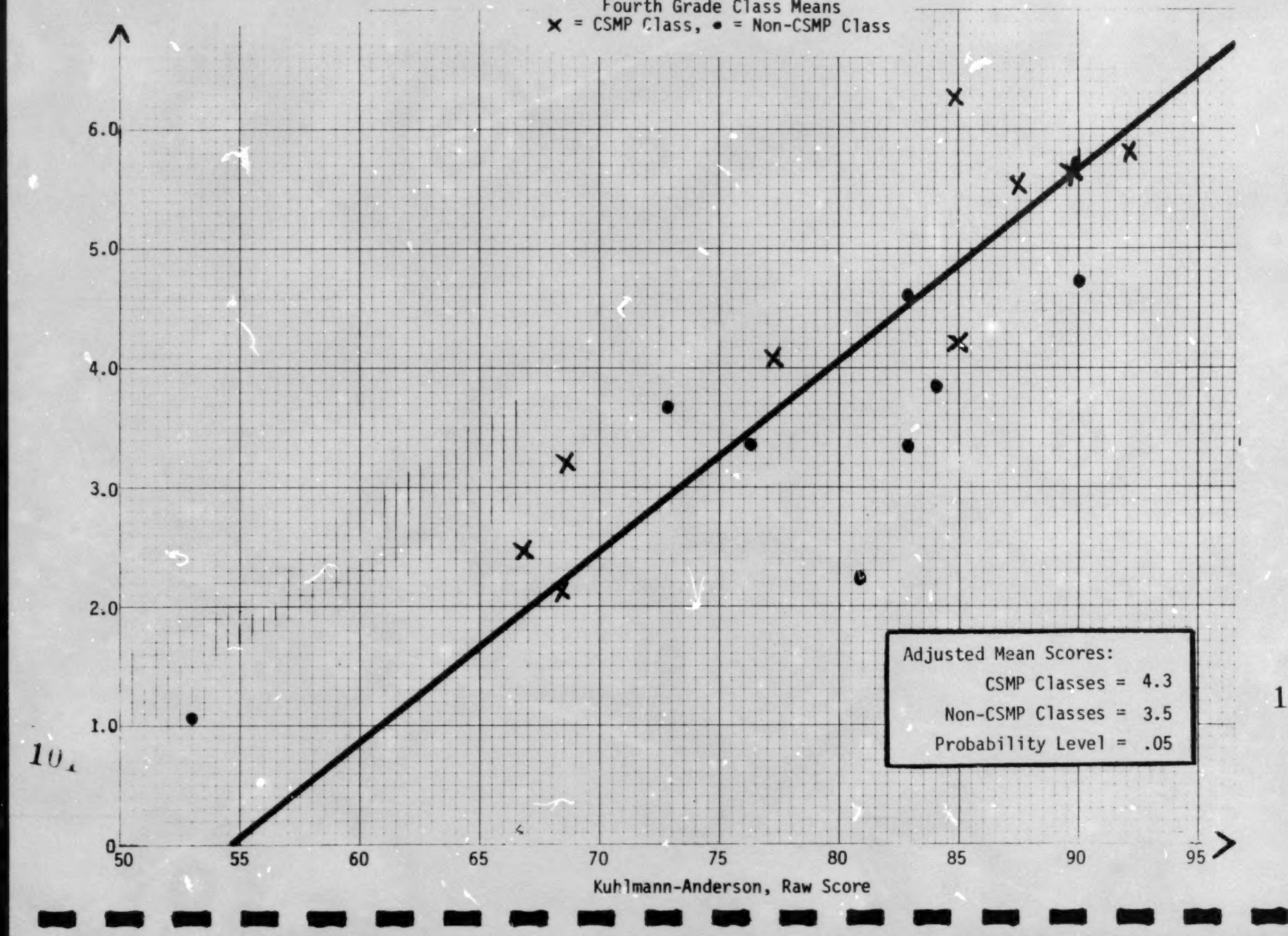
Mean By Ability Level:

	1	2	3	4	All	sd
CSMP	1.9	3.9	5.3	6.4	4.4	2.5
Non-CSMP	1.1	3.2	3.3	5.5	3.3	2.5

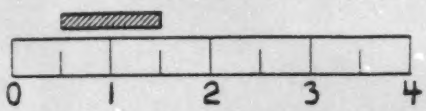
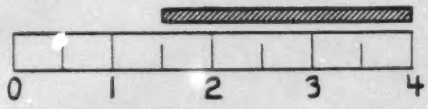
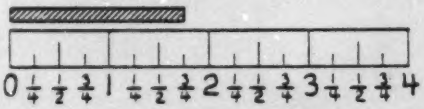
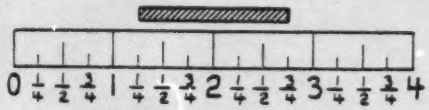
Note:

1. The first 4 items had to do with finding fractional parts of whole numbers and were given on last year's MANS. CSMP students did far better on these items the mean percent correct for CSMP and Non-CSMP students being 65 and 36 (versus 46 and 31 last year).
2. About 30% of the students gave an answer of "4" for the third and fourth items, in spite of warnings by the tester about how those questions were different from the first 2 items.

Scale #11: Fractions



SCALE #12 - Measuring

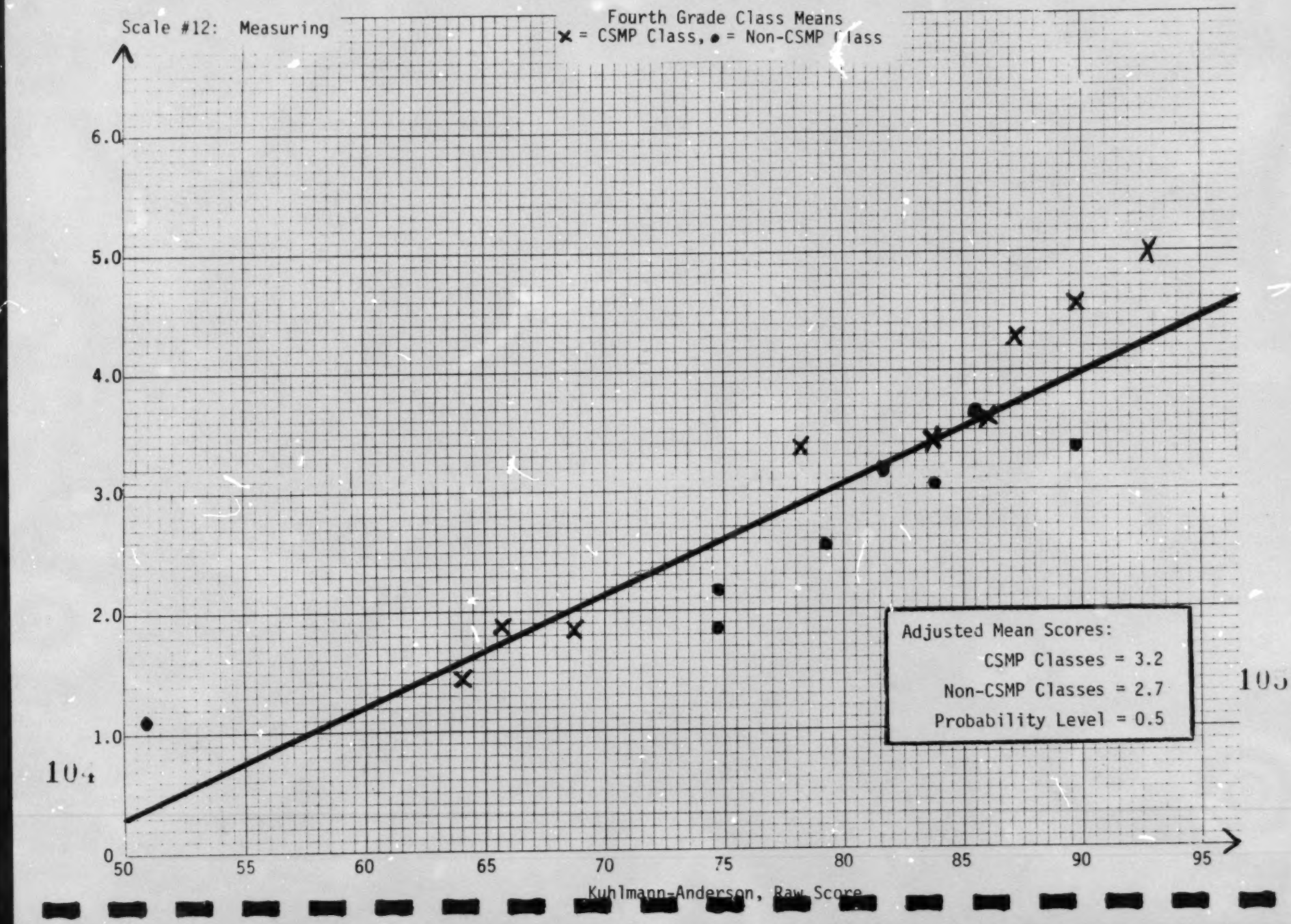
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
How Many Inches?  _____	72	67	.83	$\frac{1}{2}$, $1\frac{1}{2}$
 _____	65	65	.70	
 _____	70	70	.75	
 _____	45	29	.59	$1\frac{3}{4}$
_____	45	19	.50	
_____	24	08	.53	
Number of Students	161	141		
KR 20 Reliability Homogeneity	.75	.74		
Correlation with Kuhlmann-Anderson	.62	.65		
Mean raw score, " "	30.1	77.2		

Means by Ability Level:

	1	2	3	4	all	sd
CSMP	1.6	2.5	3.7	4.7	3.2	1.9
Non-CSMP	1.0	2.5	2.9	3.9	2.6	1.7

Note:

1. The first 4 items required measuring with a ruler marked in halves or fourths of an inch. There was little difference between CSMP and Non-CSMP except for the fourth item where for some reason, CSMP students did better. The last 2 items required measuring with a "metric" ruler (marked in tenths of an inch). CSMP students did much better on these items than did Non-CSMP students, though neither question was very well done.


Scale #12: Measuring
Fourth Grade Class Means
× = CSMP Class, ● = Non-CSMP Class
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 3.2
Non-CSMP Classes = 2.7
Probability Level = 0.5

SCALE #13 - Using Hints

Test Items*	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
$\begin{array}{r} {}^{6}\not{4},577 \\ \vdots \\ +\ \vdots \\ \hline 16,327 \end{array}$	45	38	.71	14,327
$\begin{array}{r} {}^{2}\not{4}2 \\ \vdots \\ +\ \vdots \\ \hline 574 \end{array}$	37	33	.75	594
$\begin{array}{r} 95\not{4}^{2} \\ \vdots \\ +\ \vdots \\ \hline 5,346 \end{array}$	39	22	.76	5,349
$12\not{6} \times 74 = 444$	55	46	.76	450
$2\overset{1}{\not{0}}0 \times 16 = 320$	28	17	.88	321
Number of Students	166	136		
KR 20 Reliability Homogeneity	.83	.76		
Correlation with Kuhlmann-Anderson	.57	.45		
Mean raw score, " "	79.3	77.1		

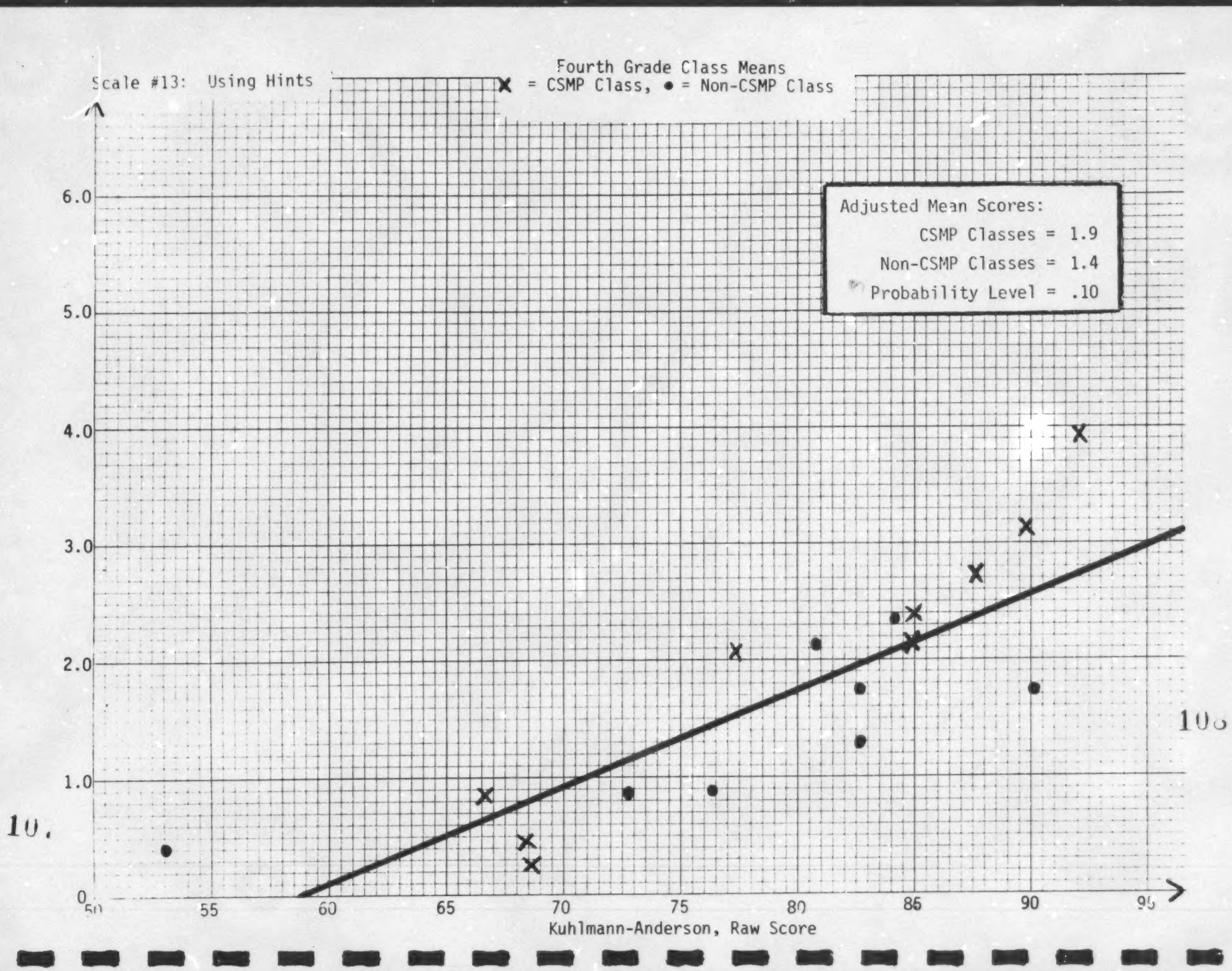
Means By Ability Level:

	1	2	3	4	All	sd
CSMP	0.5	1.3	2.0	3.7	1.8	1.8
Non-CSMP	0.4	1.0	1.4	2.5	1.3	1.6

Note:

1. This was a difficult scale for both groups of students; about 40% of the students didn't get any of the items correct. CSMP students in the 2 highest ability quartiles did better than Non-CSMP students, but there was very little difference in the lower two quartiles where neither group did particularly well.
2. For the first 3 items, about 10% of the responses were correct except that the students decreased (instead of increased) the answer by the correct amount or vice versa.

*Abbreviated here to show only the differences. See the student copy of this scale, Page A14


Scale #13: Using Hints
Fourth Grade Class Means
X = CSMP Class, • = Non-CSMP Class
Adjusted Mean Scores:
CSMP Classes = 1.9
Non-CSMP Classes = 1.4
Probability Level = .10
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score

SCALE #14a - Estimating Intervals-Addition

Test Items*	Percent Correct		Biserial
	CSMP	Non-CSMP	
Intervals: 0 10 50 100 500 1000			
185 + 97	87	84	.61
24 + 24	87	79	.38
59 + 39	88	87	.45
479 + 86	83	71	.42
279 + 165	80	77	.64
19 + 29	81	79	.64
217 + 294	67	58	.51
19 + 19 + 19	57	52	.34
Number of Students	166	136	
KR 20 Reliability Homogeneity	.63	.68	
Correlation with Kuhlmann-Anderson	.55	.63	
Mean raw score, " "	79.3	77.1	

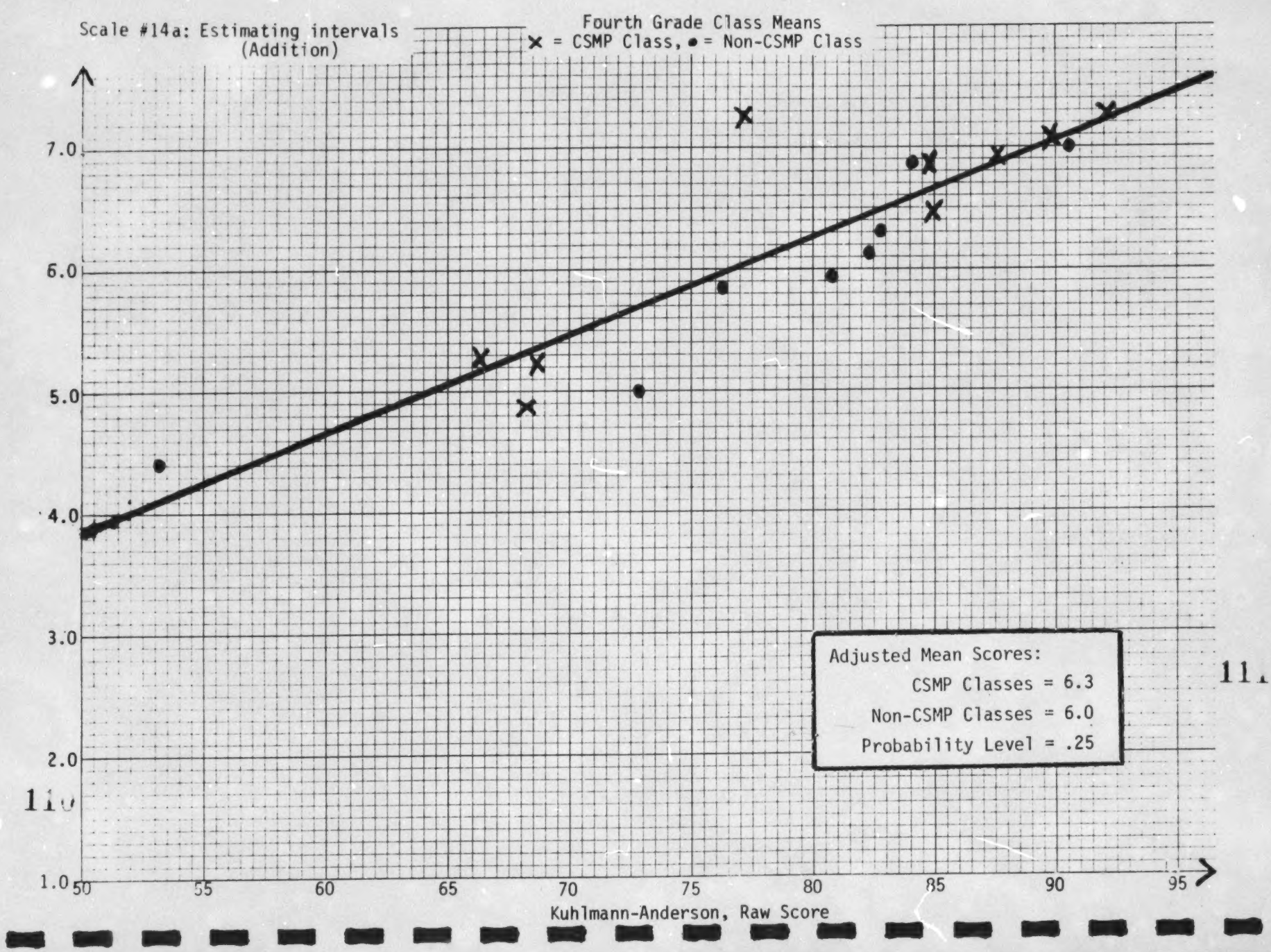
Means By Ability Level:

	1	2	3	4	All	sd
CSMP	4.9	6.0	6.8	7.3	6.3	1.7
Non-CSMP	4.1	5.6	6.8	7.2	5.9	1.9

Note:

1. See Scale #14c

*See the student copy of this scale, Page A15


Scale #14a: Estimating intervals
(Addition)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
7.0
6.0
5.0
4.0
3.0
2.0
1.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 6.3
Non-CSMP Classes = 6.0
Probability Level = .25

SCALE #14b - Estimating Intervals-Multiplication

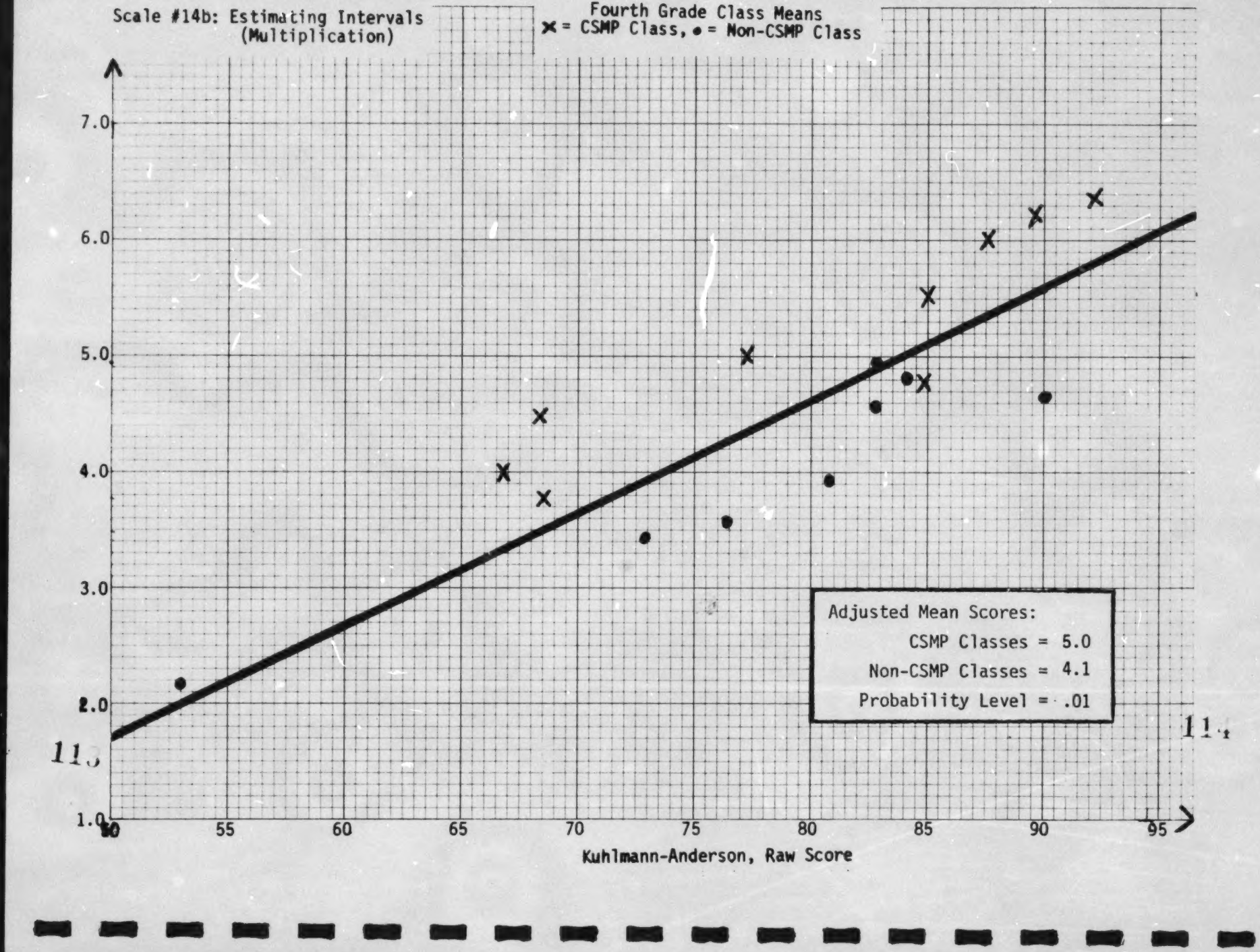
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
Intervals: 0 10 50 100 500 1000			
5 x 109	83	65	.58
2 x 19	83	75	.41
40 x 10	81	55	.61
4 x 23	70	57	.42
11 x 50	64	43	.49
2 x 49	73	68	.38
4 x 29	49	35	.44
Number of Students	166	136	
KR 20 Reliability Homogeneity	.64	.65	
Correlation with Kuhlmann-Anderson	.50	.58	
Mean raw score, " "	79.3	79.7	

Means By Ability Level:

	1	2	3	4	All	sd
CSMP	3.8	4.7	5.4	6.2	5.0	1.7
Non-CSMP	2.2	4.0	4.7	5.0	4.0	1.9

Note:

1. See Scale #14c


Scale #14b: Estimating Intervals
(Multiplication)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
7.0
6.0
5.0
4.0
3.0
2.0
1.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 5.0
Non-CSMP Classes = 4.1
Probability Level = .01

SCALE #14c - Estimating Intervals-Division

Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
Intervals: 0 10 20 50 100			
300 ÷ 4	79	71	.11
200 ÷ 9	49	49	.01
190 ÷ 10	63	50	.37
1 ÷ 2	67	49	.14
150 ÷ 13	51	47	.25
150 ÷ 100	31	30	.25
180 ÷ 21	25	18	.23
Number of Students	166	136	
KR 20 Reliability Homogeneity	.35	.32	
Correlation with Kuhlmann-Anderson	.31	.36	
Mean raw score, " "	79.3	77.1	

Means By Ability Level:

	1	2	3	4	All	sd
CSMP	2.8	3.8	3.9	4.2	3.7	1.5
Non-CSMP	2.4	3.1	3.4	3.8	3.1	1.5

Note:

1. Each of Scales 14a), 14b) and 14c) had rather low reliability measures, possibly because of guessing. This was especially true in 14c), the hardest scale, which appeared to have more "bad" answering (see below); and had the lowest correlation of all the MANS scales with the Kuhlmann-Anderson (.34).

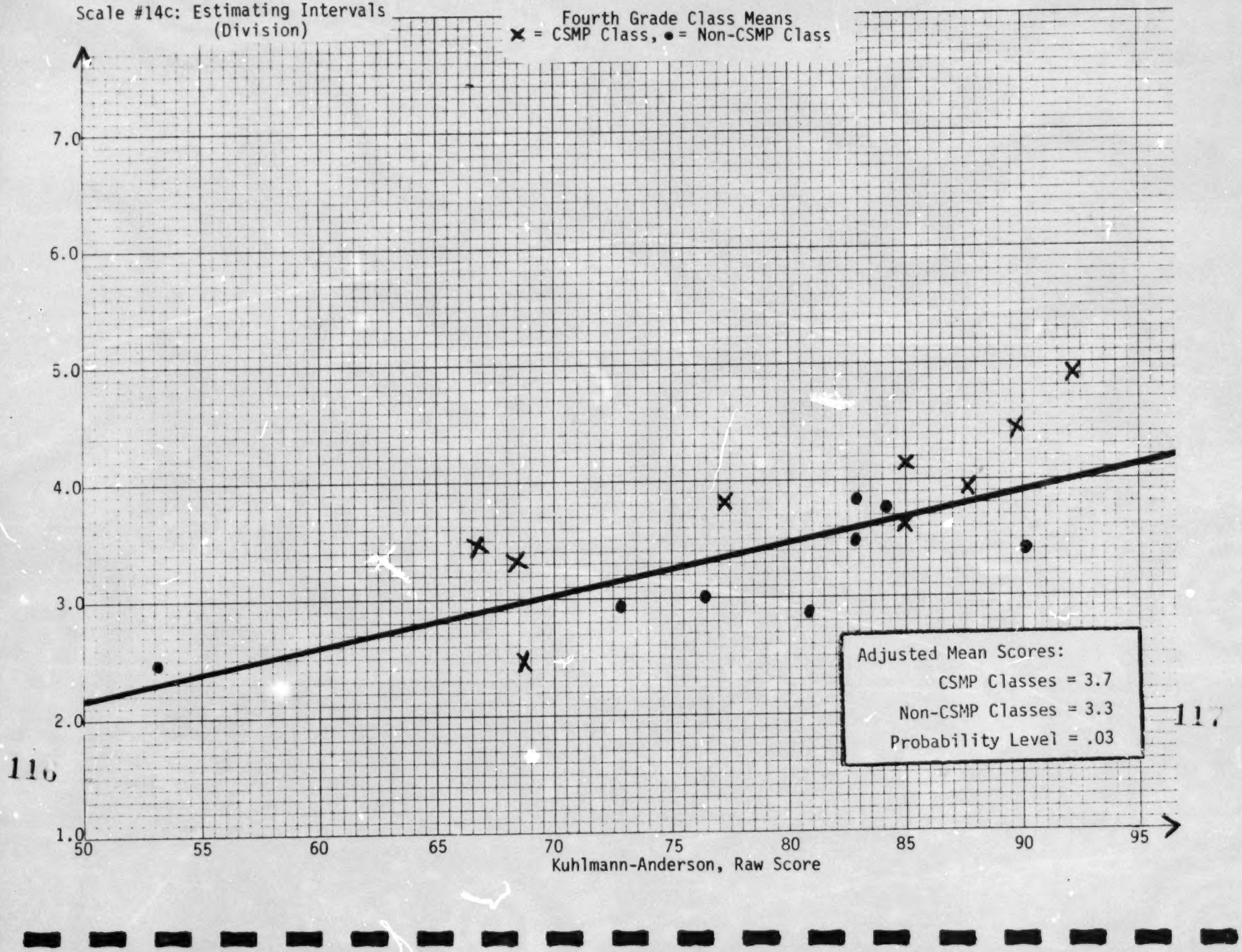
2. For each item it was possible to classify responses as correct, next best or other ("poor" response). For example, the correct interval for 2 x 49 is (50,100); the next best is (100,500); and any other response is a poor one. Then one can derive the following table:

	% Correct		% Next Best		% Poor	
	CSMP	Non-CSMP	CSMP	Non-CSMP	CSMP	Non-CSMP
Addition 14a)	79	74	16	18	3	5
Multiplication 14b)	71	57	17	20	8	19
Division 14c)	53	44	26	31	17	20

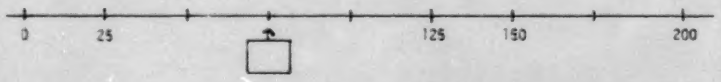
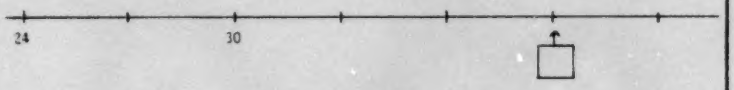
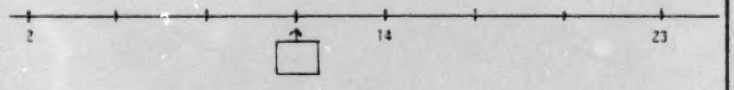
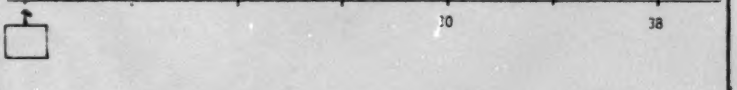
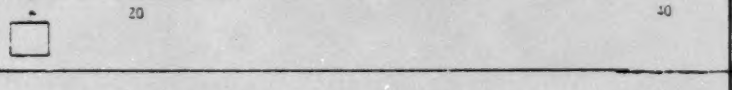
3. Students apparently had enough time for these scales; an average of 94% of the students attempted the last items in the scales.

4. The correct answer was always the most frequently chosen one, and the "next best" answer was always the second most frequently chosen with 4 exceptions: for the third multiplication item and the fifth and sixth division items, more "poor" responses were given than "next best"; and for the last division item roughly 40% of the responses were "next best", 30% "poor" and 20% correct.

5. The mean percent correct on the addition items this year was 79 for CSMP and 74 for Non-CSMP (versus 56 and 57 last year). On the multiplication items the percent correct was 71 and 57 (versus 45 and 39 last year).


Scale #14c: Estimating Intervals
(Division)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
7.0
6.0
5.0
4.0
3.0
2.0
1.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 3.7
Non-CSMP Classes = 3.3
Probability Level = .03

SCALE #15 - Labelling Number Line

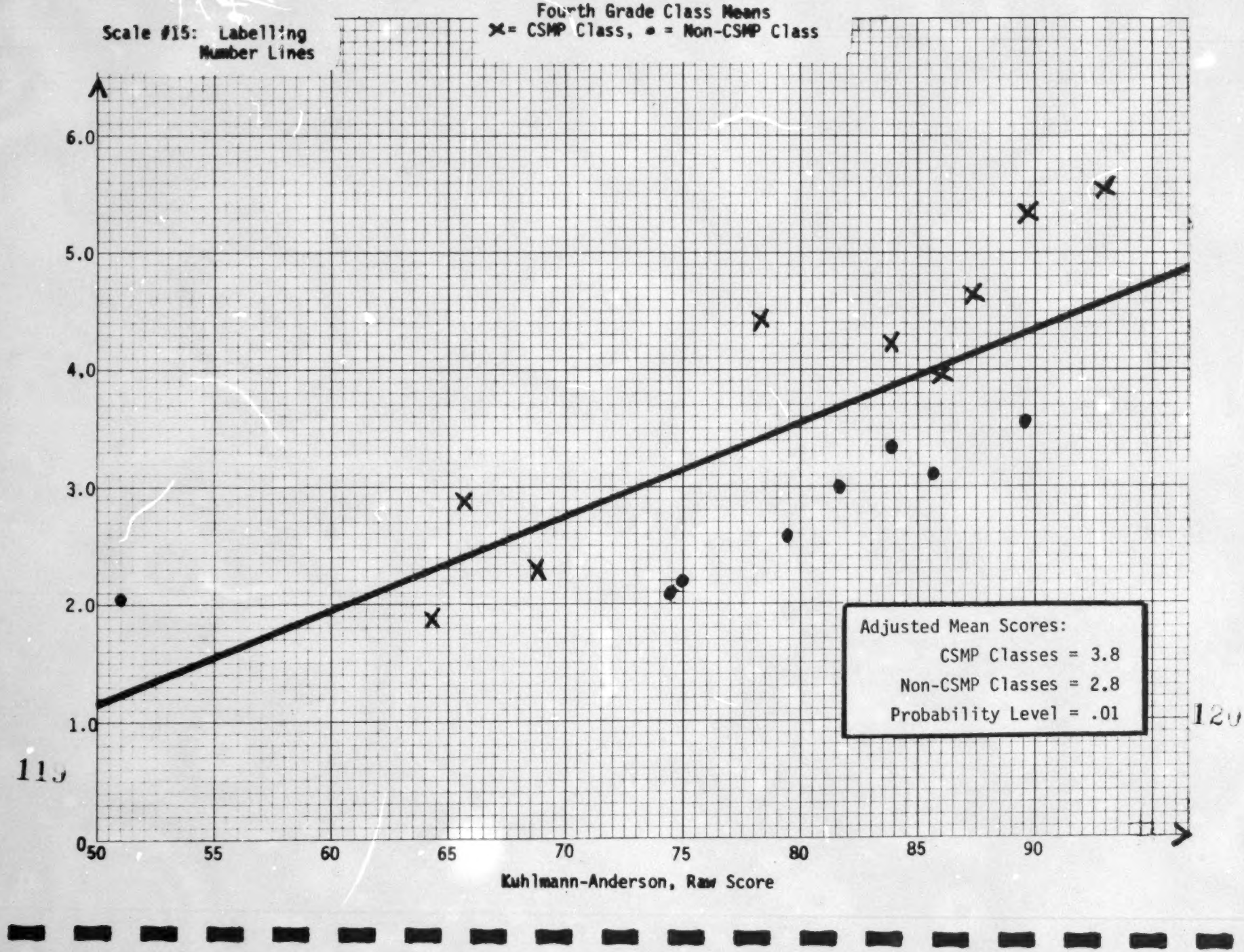
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
	76	74	.82	
	76	71	.73	50, 100
	61	41	.75	48
	66	50	.74	12
	49	16	.69	10, 22
	49	21	.68	10
Number of Students	141	161		
KR 20 Reliability Homogeneity	.83	.76		
Correlation with Kuhlmann-Anderson	.59	.55		
Mean raw score, " "	80.1	77.2		

Means By Ability Level:

	1	2	3	4	All	sd
CSMP	2.1	3.3	4.5	5.3	3.8	2.1
Non-CSMP	1.3	2.6	3.1	3.9	2.7	1.8

Note:

1. The largest differences in favor of CSMP occurred on the third, fifth and sixth items which required extrapolation outside the interval (rather than interpolation within the interval).
2. The first, second and fourth items were also administered last year. The mean percent correct on these items was 73 for CSMP students and 65 for Non-CSMP (versus 49 and 41 last year).


Fourth Grade Class Means
X= CSMP Class, • = Non-CSMP Class
Scale #15: Labelling Number Lines
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 3.8
Non-CSMP Classes = 2.8
Probability Level = .01

#SCALE #16 - Multiplying by 2, 5 or 10

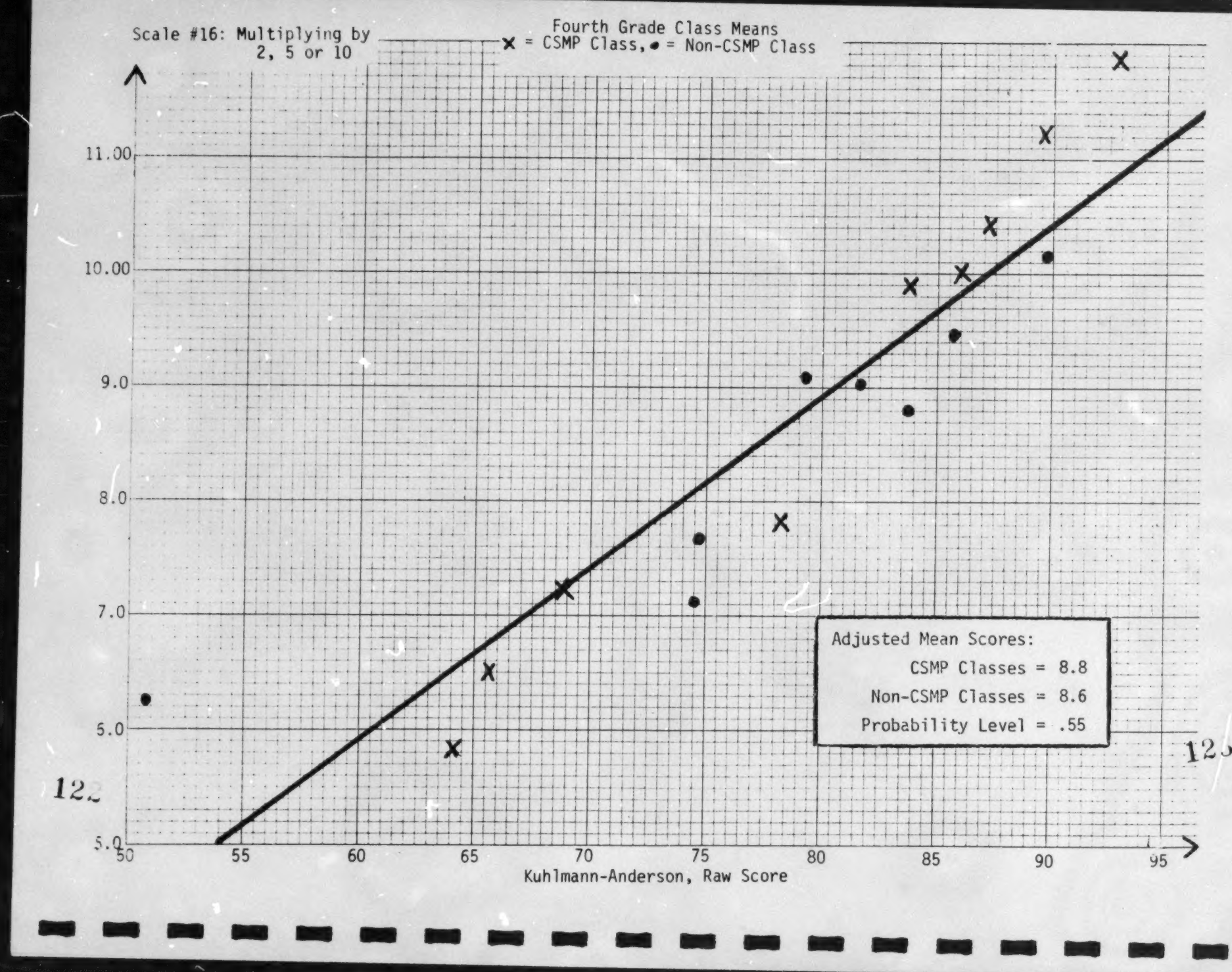
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
51 is about ___times as large as 31	84	86	.56
16 ... _____ ... 7	80	82	.61
65 ... _____ ... 12	76	67	.70
200 ... _____ ... 21	90	91	.58
100 ... _____ ... 19	43	37	.45
98 ... _____ ... 50	71	66	.88
45 ... _____ ... 8	85	74	.76
602 ... _____ ... 298	56	55	.60
503 ... _____ ... 49	76	69	.47
59 ... _____ ... 31	75	82	.75
12 ... _____ ... $2\frac{1}{2}$	49	49	.56
$20\frac{1}{2}$ ... _____ ... $10\frac{1}{2}$	71	70	.71
$5\frac{1}{2}$ ... _____ ... $\frac{1}{2}$	22	14	.31
Number of Students	161	141	
KR 20 Reliability Homogeneity	.83	.77	
Correlation with Kuhlmann-Anderson	.60	.58	
Mean raw score, " "	80.1	77.2	

Means by Ability Level:

	1	2	3	4	All	sd
CSMP	6.2	8.7	9.7	11.1	8.7	3.2
Non-CSMP	6.0	8.4	9.1	10.3	8.4	2.8

Note:

1. CSMP students did slightly better in the upper ability quartiles.
2. The first 10 items were also administered last year. The mean percent correct for CSMP and Non-CSMP was 73 and 71 respectively this year (versus 63 and 56 last year).
3. The correct answer was always the most popular choice except "100 is about___ times as large as 19", which was answered most often by "10"; and the last item where the correct answer "10" was the least popular choice.


Scale #16: Multiplying by 2, 5 or 10
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
11.00
10.00
9.0
8.0
7.0
6.0
5.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 8.8
Non-CSMP Classes = 8.6
Probability Level = .55

SCALE #17 - Which is Larger?

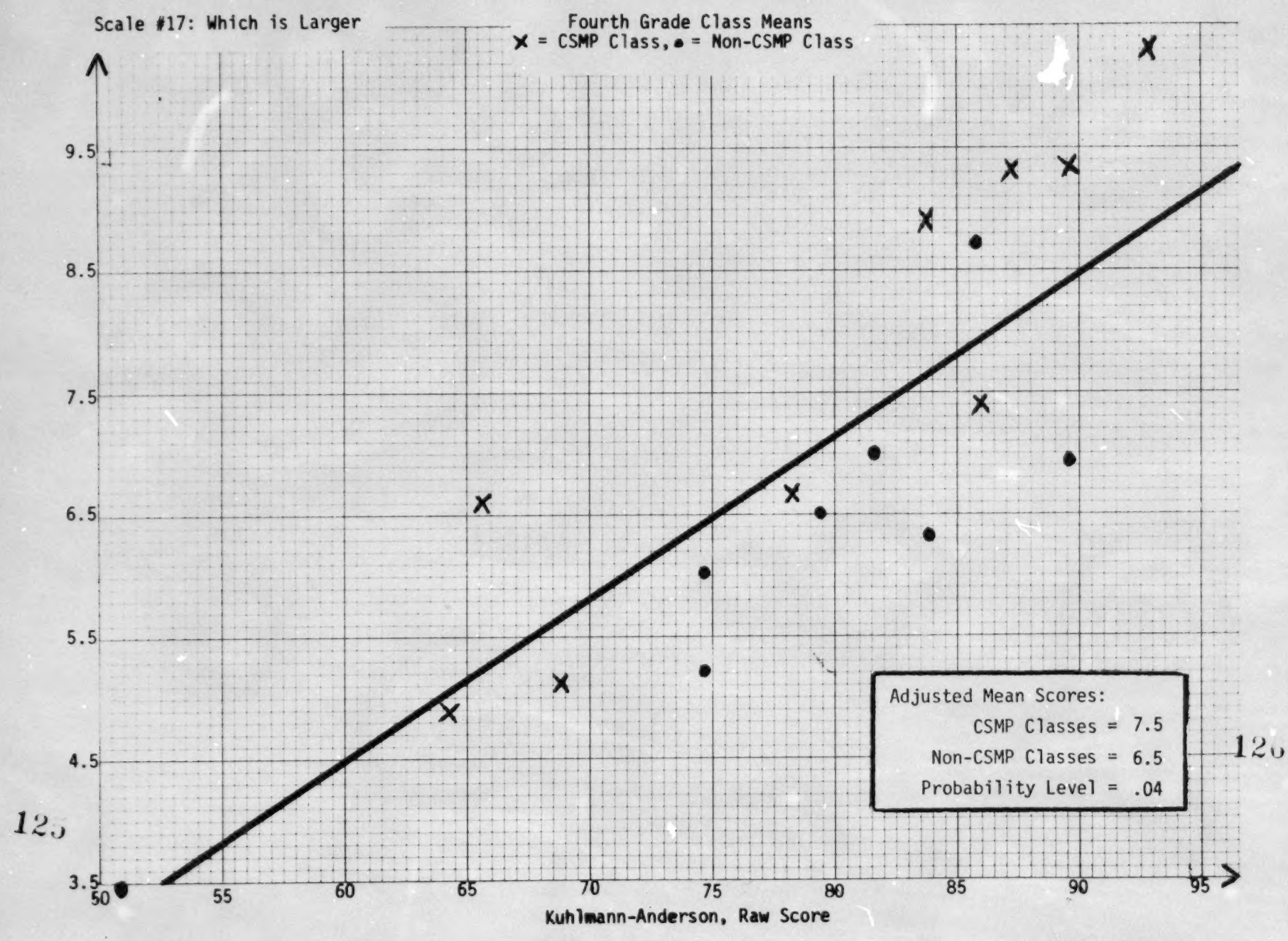
Test Items	Percent Correct		Biserial
	CSMP	Non-CSMP	
899+ 899 + 899 + 899 , 899 x 5	88	79	.53
8,000 , 8 x 16,000	92	97	.20
38 x 38 , 39 x 31	66	74	.36
3 x 162 , 4 x 160	77	69	.24
6,284 ÷ 7 , 6,284 ÷ 8	68	45	.78
599 ÷ 2 , 300	47	31	.62
800 ÷ 3 900 ÷ 31	66	46	.81
25,000 ÷ 60 , 26,000 ÷ 90	63	40	.85
$\frac{3}{8}$, $\frac{3}{4}$	44	26	.74
$\frac{1}{2}$, $\frac{2}{5}$	39	32	.73
$\frac{1}{3}$ of 19,000 , $\frac{1}{2}$ of 19,000	55	39	.71
$\frac{2}{3}$ of 500 , $\frac{1}{6}$ of 500	40	49	.39
Number of Students	161	141	
KR 20 Reliability Homogeneity	.82	.78	
Correlation with Kuhlmann-Anderson	.52	.60	
Mean raw score, " "	80.1	77.2	

Means By Ability Level:

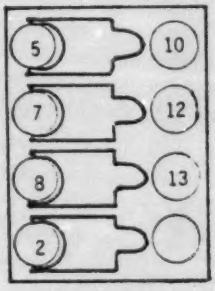
	1	2	3	4	All	sd
CSMP	5.6	6.3	8.7	9.7	7.5	3.2
Non-CSMP	3.8	5.9	7.1	8.5	6.3	2.9

Note:

1. This scale consisted of 3 sets of 4 items each, dealing with multiplication, division and fractions. The mean percent correct for CSMP and Non-CSMP students respectively was 81 and 80 (multiplication), 61 and 41 (division) and 45 and 36 (fractions).
2. Several items were answered correctly at below guessing level (i.e., by less than half the students). This was most notable in the comparison of fractions (3/8 versus 3/4, and 1/2 versus 2/5).


Scale #17: Which is Larger
Fourth Grade Class Means
X = CSMP Class, • = Non-CSMP Class
9.5
8.5
7.5
6.5
5.5
4.5
3.5
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 7.5
Non-CSMP Classes = 6.5
Probability Level = .04

SCALE #18 - Solving Number Machines

Test Items*	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
General Format: 				
(5,10) (7,12) (8,13) (2,)	88	72	.84	
(3,1) (11,9) (10,8) (15,)	85	79	.81	
(10,5) (4,2) (8,4) (6,)	83	65	.87	1
(3,12) (5,20) (10,40)(4,)	74	60	.85	
(2,6) (5,9) 10,14) (,12)	84	60	.92	
(6,3) (16,13) (8,5) (, 9)	70	54	.85	6
(6,12) (2,4) (7,14) (,8)	80	65	.91	
(9,3) (15,5) (30,10) (,2)	58	51	.79	8
Number of Students	161	141		
KR 20 Reliability Homogeneity	.86	.87		
Correlation with Kuhlmann-Anderson	.61	.66		
Mean raw score, " "	80.1	77.2		

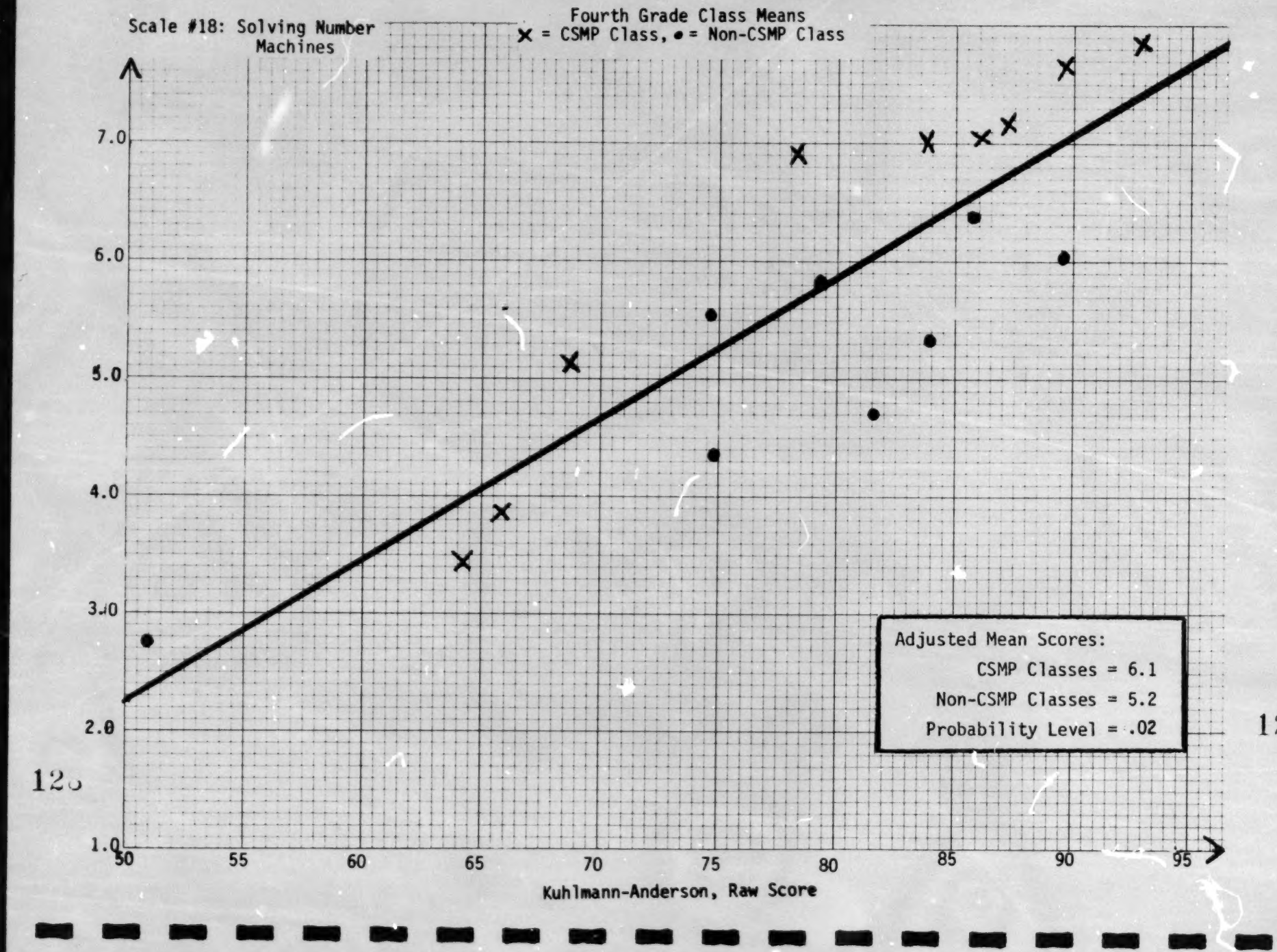
Means By Ability Level:

	1	2	3	4	All	sd
CSMP	4.8	5.9	7.2	8.1	6.2	2.3
Non-CSMP	3.3	4.8	5.3	5.8	5.1	2.7

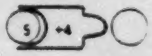
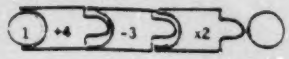
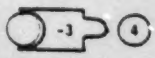
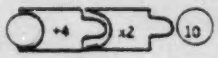
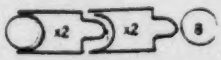
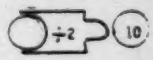
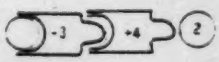
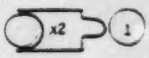
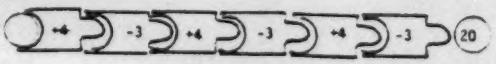
Note:

1. This scale was easy for many students; 46% of the CSMP students and 32% of the Non-CSMP students got every item correct and overall about 70% of the students got at least half the items correct.
2. This scale was also administered last year. The mean percent correct this year was 78 for CSMP students and 64 for Non-CSMP students (versus 55 and 43 last year).

*The student copy of this scale consisted of 2 sample problems and 8 test items, all within a box and in the general format shown here.


Scale #18: Solving Number Machines
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
7.0
6.0
5.0
4.0
3.0
2.0
1.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 6.1
Non-CSMP Classes = 5.2
Probability Level = .02

SCALE #19 - Using Number Machines

Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
	99	96	.80	
	91	86	.65	
	93	91	.69	
	86	69	.88	
	86	67	.83	4
	69	32	.59	5 (39%)
	47	13	.28	2
	41	02	.60	1 (34%)
	39	19	.51	16
Number of Students	161	141		
KR 20 Reliability Homogeneity	.75	.70		
Correlation with Kuhlmann-Anderson	.67	.58		
Mean raw score, " "	80.1	77.2		

Means By Ability Level:

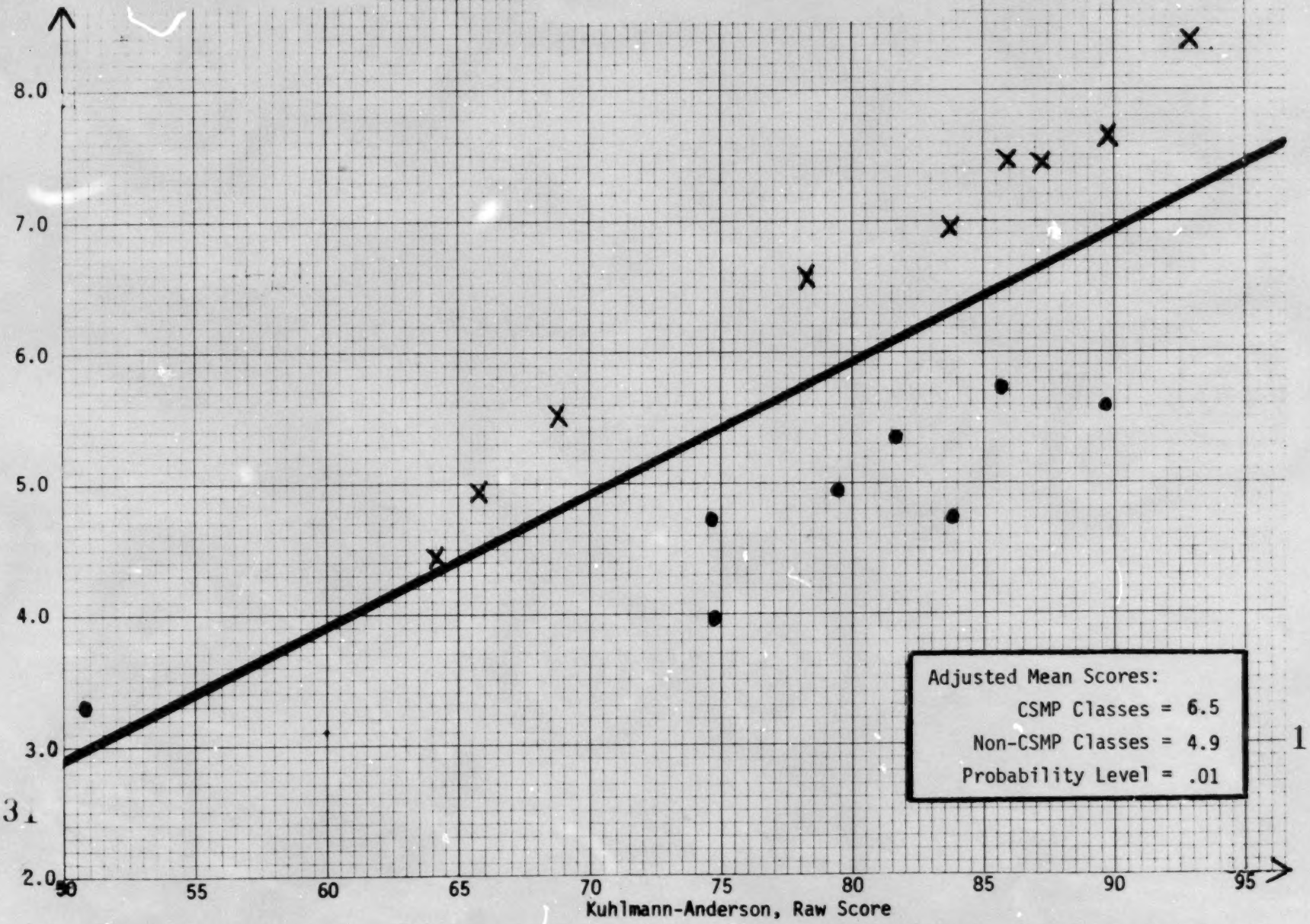
	1	2	3	4	All	sd
CSMP	4.2	6.0	6.7	7.8	6.5	2.0
Non-CSMP	2.6	4.7	5.8	7.3	4.8	1.8

Note:

1. The largest differences in favor of CSMP appear to be at the lowest ability levels (i.e., in the first two quartiles).
2. The first 5 items were also administered last year. The mean percent correct this year on these items was 91 for CSMP students and 78 for Non-CSMP (versus 70 and 60 last year).
3. The last 4 items produced huge differences in favor of CSMP. Each of these items is tricky in one way or another. The sixth item was done "backwards" by almost half of the Non-CSMP students, the seventh item is much easier if one has negative numbers available, the seventh produces a fractional answer and the last one is much easier if the "+4" and "-3" functions are composed (combined). Each of the last 3 items as omitted by about 20% of the students (versus less than 4% for the other items).

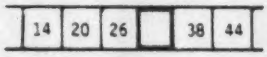
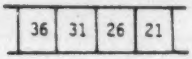
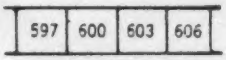
Scale #19: Using Number Machines

Fourth Grade Class Means
X = CSMP Class, • = Non-CSMP Class



Adjusted Mean Scores:
CSMP Classes = 6.5
Non-CSMP Classes = 4.9
Probability Level = .01

SCALE #20a - Boxes (1)

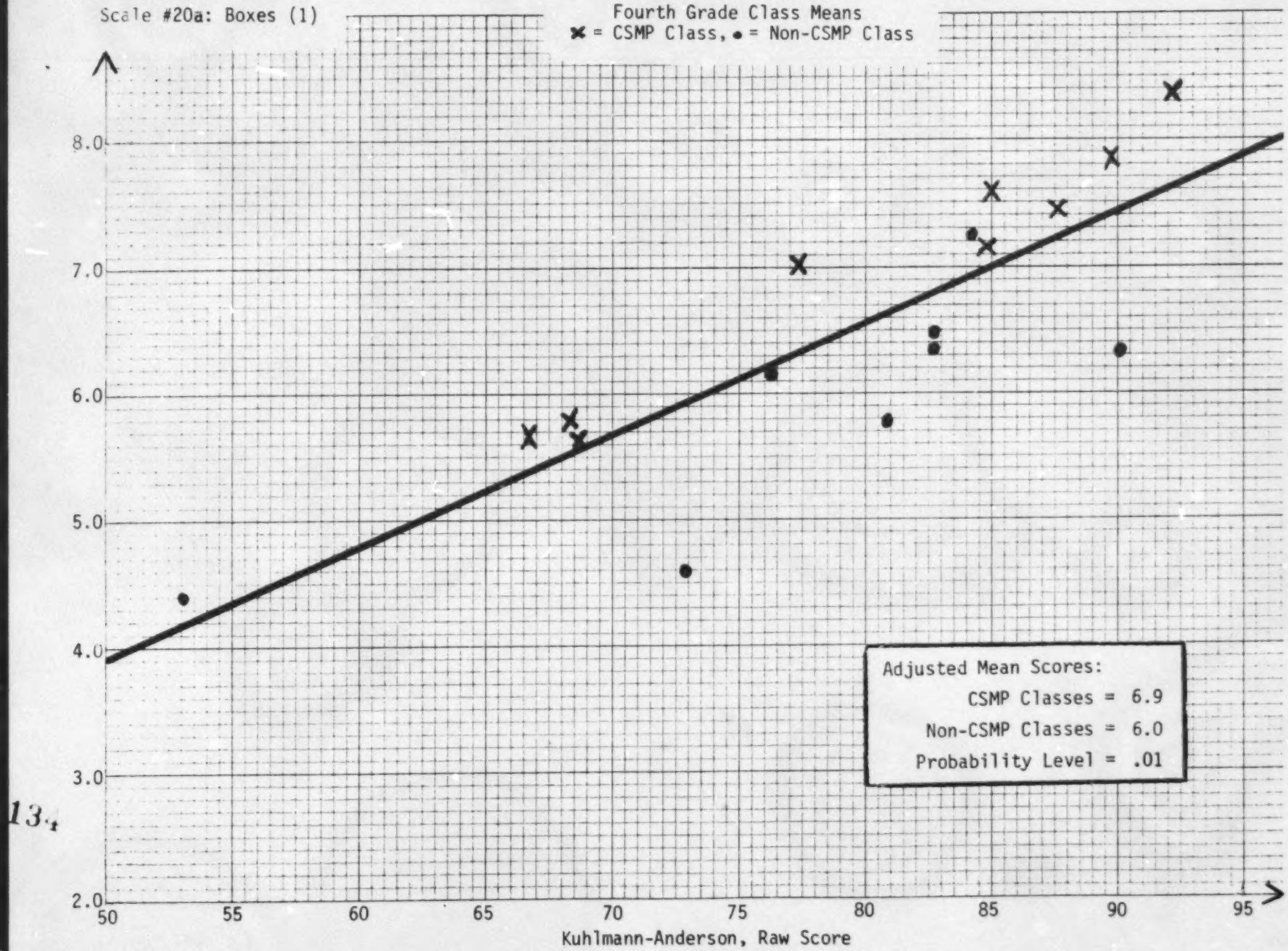
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation
 Counting by 6's	92	86	.42
Counting by 2's	79	66	.47
Will 10 be in any of the boxes? _____	76	63	.36
 Counting by 5's			
Will 51 be in any of the boxes? _____	75	64	.75
Will 86 be in any of the boxes? _____	75	55	.67
Will 980 be in any of the boxes? _____	78	66	.61
 Counting by 3's			
Will 614 be in any of the boxes? ____	84	76	.60
Will 567 be in any of the boxes? ____	63	49	.38
Will 309 be in any of the boxes? ____	64	59	.36
Number of Students	166	136	
KR 20 Reliability Homogeneity	.68	.72	
Correlation with Kuhlmann-Anderson	.49	.65	
Mean raw score, " "	79.3	77.1	

Means By Ability Level:

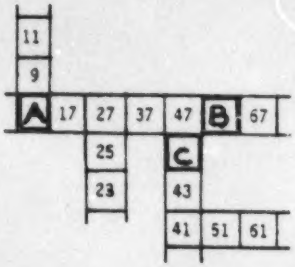
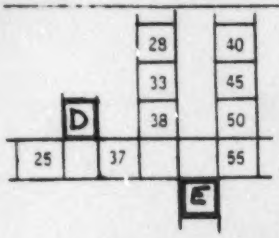
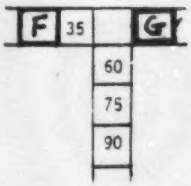
	1	2	3	4	All	sd
CSMP	5.3	6.9	7.3	7.8	6.8	2.0
Non-CSMP	3.7	5.7	7.0	7.3	5.8	2.3

Note:

1. The differences in favor of CSMP students were largest at the lowest ability level.


Scale #20a: Boxes (1)
Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
8.0
7.0
6.0
5.0
4.0
3.0
2.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 6.9
Non-CSMP Classes = 6.0
Probability Level = .01

SCALE #20b - Boxes (2)

Test Items*	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial
			
Box A	85	79	.64
Box B	93	91	.69
Box C	85	78	.68
			
Box D	34	32	.45
Box E	49	42	.56
			
Box F	54	49	.87
Box G	56	46	.93
Counting Across by?	56	46	.80
Counting Up and Down by?	70	49	.87
Number of Students	166	136	
KR 20 Reliability Homogeneity	.83	.83	
Correlation with Kuhlmann-Anderson	.56	.62	
Mean raw score, " "	79.3	77.1	

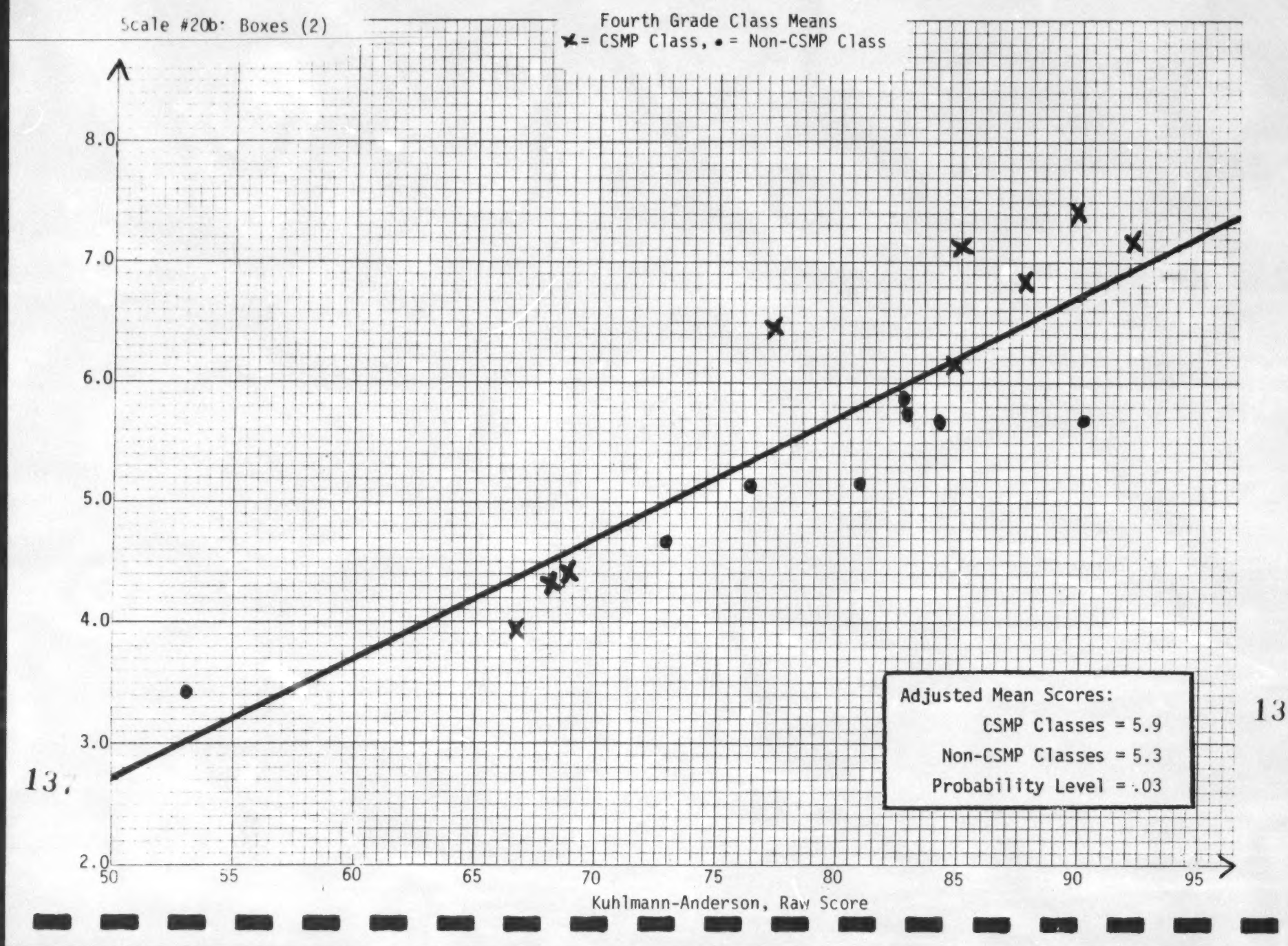
Ability By Ability Level:

	1	2	3	4	All	sd
CSMP	4.0	5.2	6.8	7.3	5.8	2.6
Non-CSMP	2.9	4.8	6.0	7.0	5.1	2.7

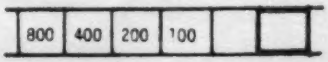
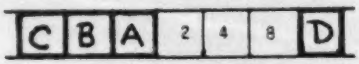
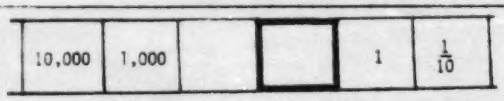
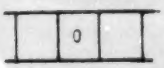
Note:

1. The first set of boxes proved to be far easier than the other two. About 10% of the students did not give any responses to the last set of boxes.

*On the student copy, the lettered Boxes were left empty for the student to fill in.


Scale #20b: Boxes (2)
Fourth Grade Class Means
✕ = CSMP Class, • = Non-CSMP Class
8.0
7.0
6.0
5.0
4.0
3.0
2.0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 5.9
Non-CSMP Classes = 5.3
Probability Level = .03

SCALE #20c - Boxes(3)

Test Items*	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation
Multiplying By 	70	37	.70
			
Box A	72	33	.79
Box B	39	07	.90
Box C	20	03	.91
Box D	69	40	.73
	54	26	.52
Counting By 48 will be in one of these boxes etc.  What could we be counting by? _____ (Give as many answers as you can.)	25	09	.42
Number of Students	166	136	
KR 20 Reliability Homogeneity	.81	.68	
Correlation with Kuhlmann-Anderson	.66	.44	
Mean raw score, " "	79.3	77.1	

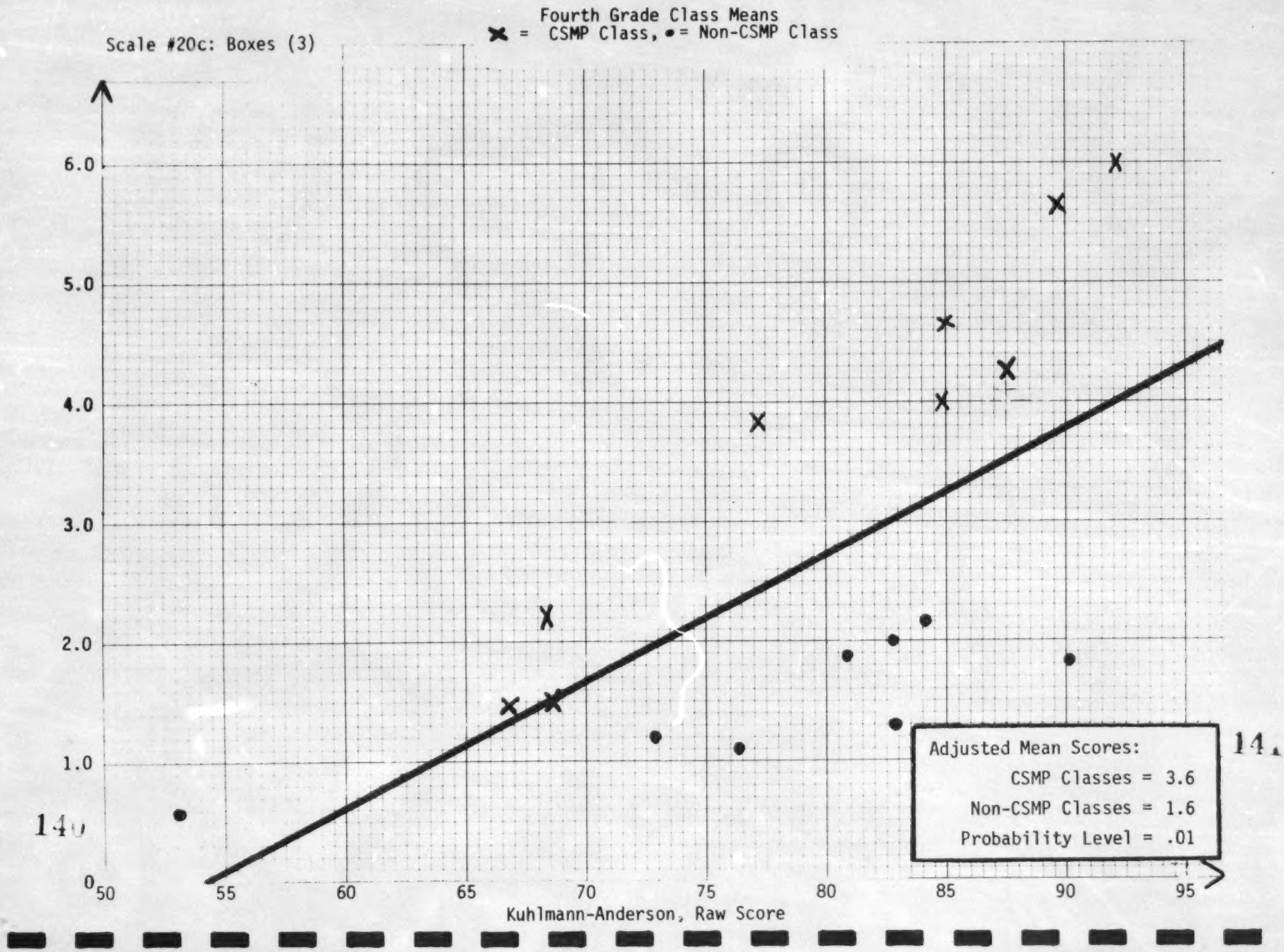
Mean Ability Score:

	1	2	3	4	All	sd
CSMP	1.5	2.9	4.4	5.4	3.5	2.2
Non-CSMP	0.6	1.3	1.5	2.5	1.5	1.6

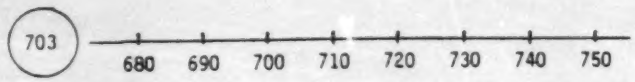
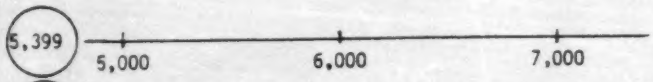
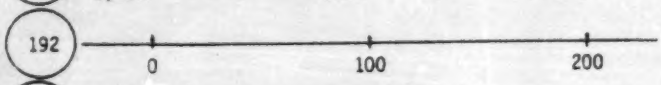
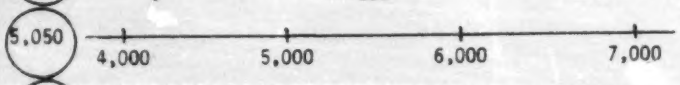
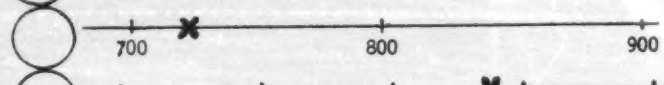
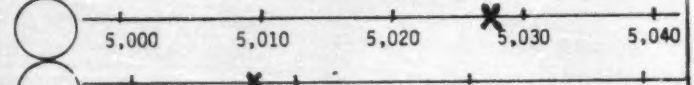
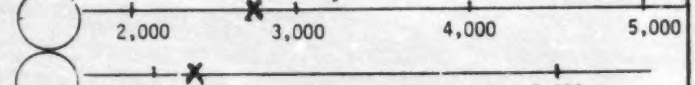
Note:

1. This scale was quite difficult for Non-CSMP students. Only 10% got more than 3 of the 7 items correct (versus 58% for CSMP).
2. In the last item there were 10 possible correct answers (excluding fractions which were accepted, though very rare). The item was counted as correct if 5 or more were given and this was done by 25% of the CSMP students and 9% of the Non-CSMP students. The percentage giving: 2-4 correct answers was 29 for CSMP and 21 for Non-CSMP; 1 correct answer was 17(34); and no correct answer was 36(28).

*On the student copy, the lettered Boxes were left empty for the student to fill in.


Fourth Grade Class Means
× = CSMP Class, • = Non-CSMP Class
Scale #20c: Boxes (3)
6.0
5.0
4.0
3.0
2.0
1.0
0
50
55
60
65
70
75
80
85
90
95
Kuhlmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 3.6
Non-CSMP Classes = 1.6
Probability Level = .01

SCALE #21 - Place Value

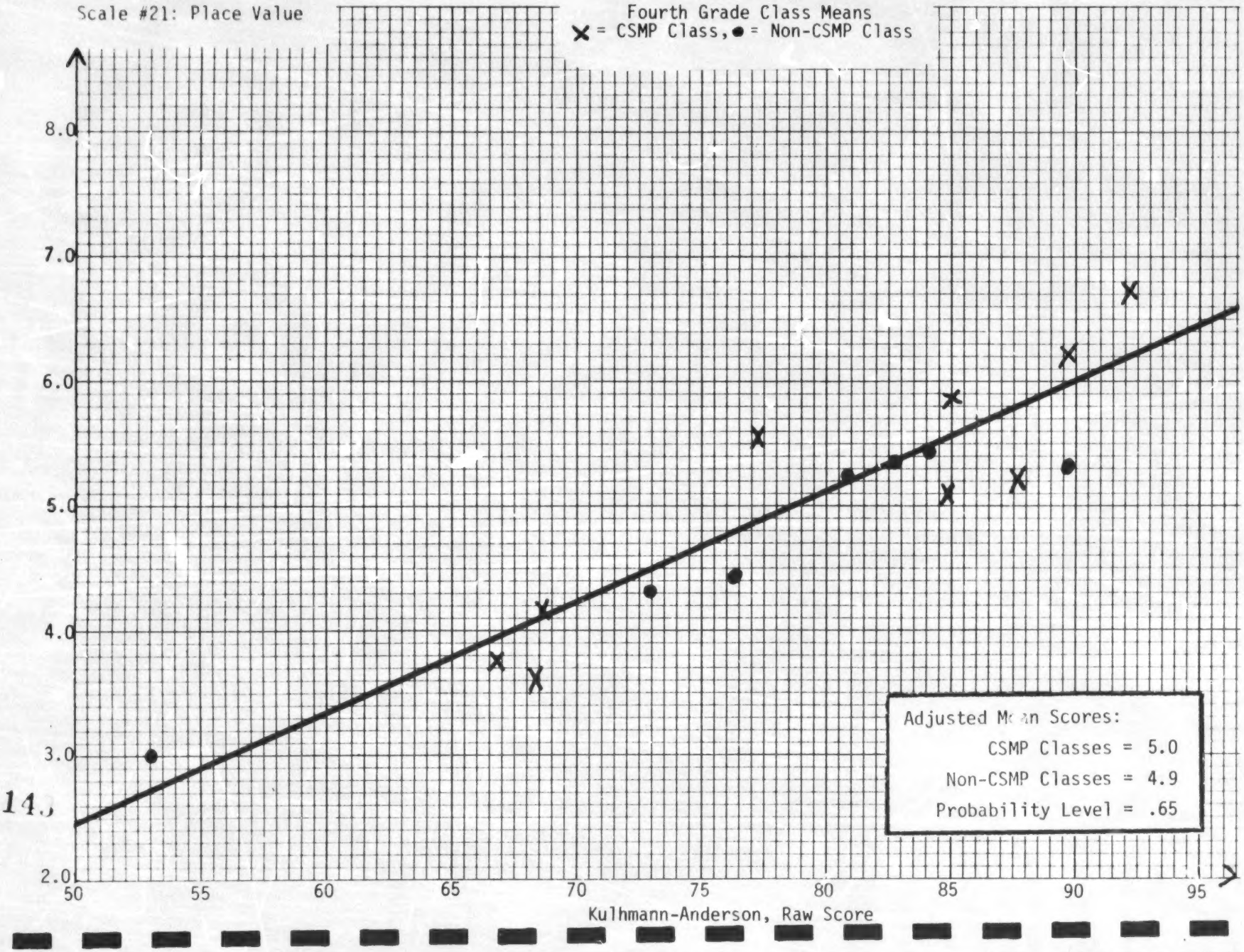
Test Items	Percent Correct CSMP	Percent Correct Non-CSMP	Biserial Correlation	Common Errors
	93	93	.33	
	84	92	.45	(5,600-6,000)
	84	73	.54	(100-140)
	34	26	.58	(5,200-5,700)-63%
	72	66	.64	(700-709)-(22%)
	50	41	.41	<5,026-(26%) (5,029-5,200)-(20%)
	20	29	.60	<2,100 -(47%) (2,900-4,000)-(14%)
	58	58	.52	(6,000-6,009)
Number of Students	166	136		
KR 20 Reliability Homogeneity	.66	.69		
Correlation with Kuhlmann-Anderson	.58	.57		
Mean raw score, " "	79.3	77.1		

Means By Ability Level:

	1	2	3	4	All	sd
CSMP	3.5	4.9	5.4	6.0	4.9	1.8
Non-CSMP	3.0	4.7	5.3	6.3	4.8	1.9

Note:

1. For each item a range was determined within which a response was counted as correct. For example, in the fourth item the "x" was counted correct if it was placed anywhere between the points corresponding to 5,000 and 5,200. Thus the scale was not intended to test how well students could estimate distances, but whether or not they attended to the relevant position or place value (ones, tens, hundreds or thousands).


Scale #21: Place Value
Fourth Grade Class Means
× = CSMP Class, ● = Non-CSMP Class
8.0
7.0
6.0
5.0
4.0
3.0
2.0
50
55
60
65
70
75
80
85
90
95
Kulhmann-Anderson, Raw Score
Adjusted Mean Scores:
CSMP Classes = 5.0
Non-CSMP Classes = 4.9
Probability Level = .65